Robustness and the Internet:
Theoretical Foundations

John C. Doyle Steven H. Low Jean M. Carlson Fernando Paganini
Glenn Vinnicombe Walter Willinger Pablo Parrilo *

Abstract

While control and communications theory have played a crucial role throughout in designing aspects of the Internet, a unified and integrated theory of the Internet as a whole has only recently become a practical and achievable research objective. Dramatic progress has been made recently in analytical results that provide for the first time a nascent but promising foundation for a rigorous and coherent mathematical theory underpinning Internet technology. This new theory addresses directly the performance and robustness of both the "horizontal" decentralized and asynchronous nature of control in TCP/IP as well as the "vertical" separation into the layers of the TCP/IP protocol stack from application down to the link layer. These results generalize notions of source and channel coding from information theory as well as decentralized versions of robust control. The new theoretical insights gained about the Internet also combine with our understanding of its origins and evolution to provide a rich source of ideas about complex systems in general. Most surprisingly, our deepening understanding from genomics and molecular biology has revealed that at the network and protocol level, cells and organisms are strikingly similar to technological networks, despite having completely different material substrates, evolution, and development/construction.

1 Introduction

Many popular technological visions emphasize ubiquitous control, communications, and computing, with systems requiring high levels of not only autonomy and adaptation, but also evolvability, scalability, and verifiability. With current technology these are profoundly incompatible objectives, and both biology and nanotechnology create additional novel multiscale challenges. A rigorous, practical, and unified theoretical framework will be essential for this vision, but until recently, has proven stubbornly elusive. Two of the great abstractions of science and technology have been the separation, in both theory and applications, of (1) controls, communications, and computing from each other, and (2) the systems level from its underlying physical substrate. These separations have facilitated massively parallel, wildly successful, explosive growth in both mathematical theory and technology, but left many fundamental problems unresolved and resulted in a poor foundation for future systems of systems in which these elements must be integrated.

As discussed in [37], much of the success of the Internet has been a result of adhering more or less faithfully over time to a set of fundamental network design principles adopted by the the early builders of the Internet (e.g., layering, fate-sharing, end-to-end). In this sense, these "principles" constitute a modest "theory" of the Internet. This theory is brilliant in the choices that were made, but it is also shallow in addressing only limited and high-level aspects of the full Internet protocol design problem. Seeking decomposition of hard problems into simpler subproblems will always be an important design strategy and is at the heart of protocols' role, but an integrated and unified theory is required to do this in a rigorous and robust manner. Nevertheless, it is pedagogically useful in introducing new theoretical concepts to decompose the key ideas in a way that both reflects the Internet protocol stack itself and builds on and makes analogies with familiar ideas. In particular, Internet researchers have drawn on all aspects of engineering systems theory, including information theory (source and channel separation, data compression, error correction codes, rate distortion), control theory (feedback, stability, robustness), optimization (duality), dynamical systems (bifurcation, attractors), and computational complexity (P-NP-coNP). Each offers some perspective on the Internet, but no one view has so far provided a foundation for the protocol suite as a whole, and in particular the

* J. Doyle and S. Low are with Caltech, Pasadena; J. Carlson is with UC Santa Barbara; F. Paganini is with UCLA, Los Angeles; G. Vinnicombe is with University of Cambridge, UK; W. Willinger is with AT&T Labs–Research, Florham Park; and P. Parrilo is with the ETH, Zürich, Switzerland.

horizontal, distributed, asynchronous and dynamic nature of control at each layer, nor the vertical interaction of the various layers of the protocol stack. In the following, we give a brief review of these myriad contributions and will sketch the new directions by making analogies with more familiar ideas.

2 The information theory analogy

A familiar analogy with information theory is that the overall network resource allocation problem, from server-client interactions through congestion control and routing to network provisioning, can be decomposed or separated into the TCP/IP protocol stack. Conventional information theory does not address this problem because of the intrinsic role that dynamics and feedback play in TCP/IP. Fortunately, a source/channel mice/elephant picture of Web and Internet traffic is emerging that begins to address these issues. New results not only complete the explanation of existing "self-similar" Internet traffic and its connection with application traffic, but are a promising starting point for a more complete source/channel coding theory analogous to that from Shannon information theory for conventional communication problems, though necessarily differing greatly in detail. Here the issue is the "vertical" separation problem: the optimality of separating the application level source vs. the TCP/IP level "channel" of the protocol stack, and then the optimal "coding" of the source and channel. The most familiar idea is that the protocol separation is, at best, optimal only in an asymptotic sense, where the asymptotes here involve large network load and capacity, as opposed to Shannon's large block size, and even in these asymptotic regimes there may be unavoidable suboptimalities. The source coding aspect is the most well-developed, with a variety of results, and models and theories of various levels of detail.

Many central aspects of internetworking that might seem related to source and channel coding problems in networks have received almost no theoretical treatment. For example, if the Web sites and clients browsing them are viewed collectively as a single aggregate "source," then this source involves both feedback and geometry as users interactively navigate hyperlinked content. While coding theory is relevant to file compression, the geometric and dynamic feedback aspects are less familiar. Furthermore, the "channel" losses experienced by this source are primarily due to congestion caused by traffic generated by the source itself. This traffic has long range correlations and is self-similar on many time scales [17, 31, 35], which in turn can be traced to heavy-tailed file distributions in source traffic being streamed out on the net [36, 7]. These discoveries have inspired recent but extensive research in the modeling of network traffic statistics, their relationship to network protocols and the (often huge) impact on network performance. Despite these efforts, the full implications of this research have yet to be understood or exploited, and only recently has there emerged a coherent coding and control theoretic treatment of the geometric and dynamic aspects of Web and other application traffic.

3 The heavy-tailed source vs. self-similar channel traffic characteristics

Heavy-tailed source behavior and the self-similar nature of aggregate network traffic initially frustrated mainstream theorists, because both characteristics violate standard assumptions in information and queueing theory. The strongly heavy-tailed nature of traffic at the source level and nearly self-similar characteristic of aggregate traffic at the link level are quite unlike the traditionally assumed Poisson traffic models. Real network traffic exhibits long-range dependence and high burstiness over a wide range of time scales. While most files ("mice") have few packets, most packets are in large files ("elephants"). It has further been widely argued that the dominant source of this behavior is due to the heavy-tailed nature of Web and other application traffic being streamed out onto the network by TCP to create long-range correlations in packet rates. The applications naturally create bandwidth-hogging elephants and delay-sensitive mice, which coexist rather awkwardly in the current Internet. Our new treatment of this problem builds on theories from robust control [26, 32] and duality in optimization [12, 21], all with a generalized coding perspective from the HOT framework [4, 5, 8, 39] and provides a radically different view. For one, we claim that heavy-tailed source traffic must be embraced, because it is not an artifact of current applications and protocols, but is likely to be a *permanent and essential feature* of network traffic, including a majority of advanced network scenarios (the bad news?). Furthermore, we show that not only can a new theory be developed to handle heavy-tailed source traffic, but if properly exploited, heavy-tailed source behavior is in fact ideal for efficient and reliable transport over packet-switched networks (the good news!). We aim to show that Web and Internet traffic can be viewed as a (perhaps very unfamiliar) joint source and channel coding problem which can be treated systematically as a global optimization problem that is implemented in a decentralized manner.

That the heavy tailed distributions associated with Web traffic are likely to be an invariant of much of the future network traffic, regardless of the application, is one important insight to be gained from this research direction, even in its currently nascent state. We expect that the current split of most traffic into mice and elephants will persist. Most files will be mice; these files make little aggregate bandwidth demand (due to their small size), but need low latency. Most of the packets come from elephants; these files demand high average bandwidth, but tolerate varying per packet latency. For example, sensor and real-time control applications naturally code into time-critical mice with measurement updates and actuator commands, against a background of elephants which update models of the dynamical environment and network characteristics. Of course, a coherent theory to make rigorous these informal observations is far from available, and the current toy model-based results are merely suggestive and encouraging.

While the empirical evidence for this mice/elephant mix in Web and other Internet traffic has received substantial attention recently, not only has no other theoretical work been done to explain it, but in fact the implications for congestion control have been largely ignored, except for the repeated assertions that these distributions break all the standard theories. Fortunately, this type of traffic creates an excellent blend when the network is properly controlled, and we have already made significant progress in exploring the profound implications of heavy-tailed traffic for network quality of service (QoS) issues. Thus two critical properties of networks converge in a most serendipitous manner: heavy tails are both a ubiquitous and permanent feature of network traffic, and an ideal mix of traffic, if properly controlled.

4 The control theory analogy

The control theory analogy derives from a duality theory of decentralized network resource allocation. This theory integrates and unifies existing approaches to TCP, and also proposes a utility and duality framework for designing new TCP/AQM strategies. This duality theory currently addresses most completely the "horizontal" decomposition of the decentralized and asynchronous nature of TCP/AQM, but it provides the framework for the mathematics for the above "vertical" theory of the protocol stack as well. The latter addresses the source/channel separation between applications and TCP/IP, as well as the further decomposition into TCP and IP routing [34], and router and link layer network provisioning [1]. A complete separation theory for the entire protocol stack with realistic modeling assumptions will likely require years to develop, but the initial results are very encouraging.

Specifically, we have shown that one can regard TCP/AQM as a distributed primal-dual algorithm carried out over the Internet by TCP sources and routers in the form of congestion control to maximize aggregate utility subject to capacity constraints at the resources [12, 21, 15, 25, 24, 38, 22, 20]. Different TCP or AQM protocols solve the same utility maximization problem with different utility functions, using different iterative algorithms. The TCP algorithm defines the the objective function of the underlying optimization problem. The AQM algorithm ensures the complementary slackness condition of the constrained optimization and generates congestion measures (Lagrange multipliers) that solve the dual problem in equilibrium. The model implies that the equilibrium properties of networks under TCP/AQM control, such as throughput, delay, queue lengths, loss probabilities, and fairness, can be readily understood by studying the underlying optimization problem. Moreover, since the problem is a concave program, these properties can be efficiently computed numerically.

It is possible to go between utility maximization and TCP/AQM algorithms in both directions [20]. We can start with general utility functions, e.g., tailored to our applications, and then derive TCP/AQM algorithms to maximize aggregate utility, as done in, e.g., [12, 21, 23, 25, 16]. Conversely, and historically, we can design TCP/AQM algorithms and then reverse-engineer the algorithms to determine the underlying utility functions they implicitly optimize and the associated dual problem. This is the consequence of end-to-end control: as long as the end-to-end congestion measure to which the TCP algorithm reacts is the *sum* of the constituent link congestion measures, such an interpretation is valid.

A critical recent development concerns the robustness and stability of the dynamics of TCP/AQM. These results clarify the dynamic (including instability) properties of existing protocols [9, 19], and have led to new designs which are provably, robustly scalable for large network size, capacity, and delay [26, 32, 14, 13, 33, 28, 6, 27]. The central technical issue here is the stability of a fully decentralized and asynchronous feedback control system. Here it is insufficient to take asymptotic limits, and one needs to prove robustness to arbitrary topologies and delays. The main insight from this series of work is to scale down source responses with their own round trip times and scale down link responses with their own capacities, in order to keep the gain over the feedback loop under control. It turns out that the required scaling with respect to capacity [26] is already built-in at the links if sources react to queueing delay, as done in [2, 22, 6, 11], as opposed to packet loss. Delay as a congestion signal is also observable at the source, making it possible to stabilize a TCP/IP network by only modifying the hosts but not the routers [10, 11].

The duality model clarifies that any TCP/AQM algorithms can be interpreted as maximizing aggregate utility in equilibrium and the stability theory explains how to design these algorithms so that the equilibrium is stable, under the assumption that the routing is given and fixed. Can TCP–AQM/IP, with minimum cost routing, jointly solve the utility maximization over both source rates and their routes? The dual problem of utility maximization over both source rates and routing has an appealing structure that makes it solvable by minimum cost routing using congestion prices generated by TCP–AQM as link costs. This raises the tantalizing possibility that TCP–AQM/IP may indeed turn out to maximize utility with proper choice of link costs. We show in [34] however that the primal problem is NP-hard, and hence cannot be solved by minimum cost routing unless P=NP. This prompts the questions: How well can IP solve the utility maximization approximately? In particular, what is the effect of the choice of link cost on maximum utility and on routing stability? For the special case of a ring network with a common destination, we show that there is an inevitable tradeoff between utility maximization and routing stability. Specifically, link costs and minimum cost routing form a feedback system. This system is unstable when link costs are pure prices. It can be stabilized by adding a static component to the link cost. The loss in utility however increases with the weight on the static component. Hence, while stability requires a small weight on prices, utility maximization favors a large weight.

While duality theory plays again a key role in formulating and studying TCP–AQM/IP-related problems, the implicit assumption is that the underlying physical network infrastructure/network design is given and fixed. Can the above-described nascent mathematical theory of the Internet be extended to layers below IP by incorporating the decentralized mechanisms and forces at work for link layer network provisioning? Note that the ability to make detailed measurements at the relevant layers and an in-depth knowledge of how the individual parts work and are interconnected make it possible to unambiguously diagnose and "reverse engineer" any such proposed theory and allow for a clean separation between sound and specious approaches. To this end, we argue in [1] that by focusing on the causal forces at work in network design and aiming at identifying the main economic and technological constraints faced by network designers and operators, observed large-scale network behavior (e.g., hierarchical structures, statistical graph properties) can be expected to be the natural by-product of an optimization-based approach to link-layer network provisioning. In particular, we show in [18] that when imposing certain technological constraints that reflect physical hardware limitations (e.g., router capacity), the space of possible network designs with identical large-scale behavior (expressed in terms of the node degree distributions of the resulting graphs) divides into two very different groups. One group consists of the set of generic or "random" configurations, all of which have poor performance and high cost and low robustness. In contrast, the other group contains all non-generic or "highly-designed" configurations, and they are characterized by high performance, low cost, and high robustness.

5 Complexity issues

The ultimate challenge in this overall research program is creating a theoretical infrastructure for formal and algorithmic verification of the correctness and robustness of scalable network protocols and embedded software for control of distributed systems. We must ultimately be able to *prove* that systems are robust, and at multiple levels of abstraction, including that of embedded code. The two approaches traditionally most concerned with these issues are robust control and formal software verification, which until recently have been both limited in their own domains, and largely independent. New results both radically expand and extend these methods, but also offer hope for their ultimate integration. In both cases, the main conceptual objective is to guarantee that a clearly defined set of "bad behaviors" is empty. For example, in the case of robustness analysis of linear systems, that set can correspond to a particular combination of uncertain parameters producing a large performance index. In protocol verification the bad behavior can be associated, for instance, to a deadlock condition.

Under minimal assumptions, most verification problems can be shown to belong to the computational complexity class known as co-NP, and the asymmetry between the classes NP and co-NP are as important and profound as those between P and NP (unless they are the same, which is unlikely). Problems in NP always have short proofs, while those in co-NP have short refutations. In contrast, there is no reason in principle to expect co-NP proofs to be short, i.e., polynomial time verifiable. It is a remarkable fact that in many cases, concise arguments about robustness (or correctness) can be provided. The consequences and implications of this finding are not fully understood. A promising theoretical framework for the unification of robustness and verification methods is the systematic theory of convex relaxations for co-NP problems recently developed in [30, 29]. The breakthrough in Parrilo's work is that the search for *short proof certificates* can be carried out in an *algorithmic way*. This is achieved by coupling efficient optimization methods and powerful theorems in semialgebraic geometry.

Short proofs and relaxations may produce solutions that are actually better (less brittle, more scalable, evolvable, etc) than the original formulation. For instance, there are many results describing an inverse relationship between the "hardness" of a problem (or its proof length) and the distance to the set of ill-posed instances. Similarly, the IP layer was designed primarily for simplicity and robustness while performance was sacrificed. Scalability was a less explicit consideration, but one that turned out to be a very strong property of the design. Our work on robustness analysis of TCP/AQM/IP suggests that the protocol may be *provably* robust as well, again not a consideration originally. Indeed, the TCP needed to make this proof work is not the one originally designed nor is it one currently deployed. Thus designing for extreme robustness has produced a protocol that was remarkably scalable, and may evolve into one that is verifiable as well.

Questions in complex systems are typically nonlinear, nonequilibrium, uncertain, hybrid, etc., and their analysis has relied mainly on simulation. By contrast, our program focuses on rigorous analysis. One computer simulation produces one example of one time history for one set of parameters and initial conditions. Thus simulations can only ever provide counterexamples to hypotheses about the behavior of a complex system, and can never provide proofs. In technical terms, they can in principle provide satisfactory solutions to questions in NP, but not to questions in co-NP. Simulations can never prove that a given behavior or regularity is necessary and universal; they can at best show that a behavior is generic or typical. What is needed at this juncture is an effective (and scalable) method for, in essence, systematically proving robustness properties of nonlinear dynamical systems. That such a thing could be possible (especially without P=NP=co-NP in computational complexity theory) is profound and remarkable, and it is the main focus of our approach.

6 Robust yet fragile

A key aspect of complex systems that has been a main theme of HOT [4, 5, 8, 39, 3] is the relationship between robustness and fragility. It has the computational counterpart of "dual complexity implies primal fragility." Practically speaking, the "robust yet fragile" viewpoint completely changes what is possible. Organisms, ecosystems, and successful advanced technologies are highly constrained in that they are not evolved/designed arbitrarily, but necessarily in ways that are robust to uncertainties in their environment and their component parts. The ensuing complexity makes these systems robust to the uncertainties for which such complexity was selected, but also makes the resulting system potentially vulnerable to rare or unanticipated perturbations. This complexity is also largely cryptic and hidden in idealized laboratory settings and in normal operation, becoming conspicuous acutely when contributing to rare cascading failures or chronically through fragility/complexity evolutionary spirals [37]. These puzzling and paradoxical features are an ongoing source of confusion to experimentalists, clinicians, and theorists alike, and have led to a rash of specious theories both in biology and more recently about the Internet. However, these "robust yet fragile" features are neither accidental nor artificial and derive from a deep and necessary interplay between complexity and robustness, modularity, feedback, and fragility. Failure to explicitly exploit the highly structured, organized, and "robust yet fragile" nature of such systems hopelessly dooms other methods to be overwhelmed by their sheer complexity.

The HOT concept was introduced to focus attention on the robust yet fragile character of complexity. This is the most essential and common property of complex systems in biology, ecology, technology, and socio-economic systems. HOT offers a new and promising framework to study not only network problems, but also put networks in a bigger context. This will be important both with the convergence of existing communication and computing networks and their widely proposed role as a central component of vast enterprise and global networks of networks including transportation, energy, logistics, etc. Research within the HOT framework addresses many complementary aspects of the multifaceted area of networked complex systems. Issues such as robustness, scalability, verifiability and computability can now (and must) be investigated and understood within a common framework. We believe that these different requirements are not only compatible, but can be combined together in a very natural fashion. Our technologies have a priori emphasized these as separate issues, and the promise of a unified approach to simultaneously handle these critical aspects is of paramount importance. While seeking decomposition of hard problems into simpler subproblems will continue to be an important design strategy and is at the heart of protocols' role, an integrated and unified theory is required to do this in a rigorous and robust manner.

7 Conclusions and outlook

The preliminary findings discussed in this article are intended to illustrate the potential and sketch some of the ingredients of the type of mathematical theory of the Internet that we envision and that would allow for a holistic treatment of the entire TCP/IP protocol stack. In particular, our recently derived TCP modifications provide a perfect example of the practical advantages of including a strong theoretical machinery as an integral part of the design rationale. They also suggest the style of proofs that we think are possible, and are consistent with the existing methods in communications, controls, and computing. We expect to be able to systematically verify the robustness and correctness of protocols, systems, and software with realistic assumptions, and to further prove approximate optimality in extreme, asymptotic scenarios, which may be severe abstraction from reality. Note we are essentially abandoning the hope of optimally robust systems as intrinsically unobtainable for problems of practical interest. Even so, we expect the expansion of the classes of problems for which we can rigorously and algorithmically verify robustness to be slow and difficult, particularly when physical layers impose resource limitations. Nevertheless, we fully expect that our approach will provide the necessary theoretical framework for designing provably robust and evolvable future network technologies. Furthermore, our approach has already demonstrated how the ensuing new insights about the Internet combine with our understanding of its origins and evolution to provide a rich source of ideas about emergence, complexity and robustness in general.

Emergence and complexity are often used in ways that seem more intent on obfuscation than clarification. Nevertheless, it is certainly fair to say that the Internet is teeming with complexity and emergence, including power law distributions, self-similarity, and fractals. It could also be described as adaptive, self-organizing, far-from-equilibrium, nonlinear, heterogeneous, etc. These are all popular notions within the community of researchers loosely organized around such rubrics as Complex Adaptive Systems (CAS), New Science of Complexity (NSOC), Chaoplexity, and more specifically Self-Organized Criticality (SOC), Edge-of-Chaos (EOC), and New Science of Networks.[1] Recently, a rash of papers have offered explanations for emergent Internet phenomena from these perspectives. Comparisons with our approach have appeared elsewhere (e.g. [37, 8, 4, 3]) and demonstrate that these CAS/NSOC/SOC/EOC-based "explanations"—serving at best as simple null hypotheses—can be convincingly debunked and are easy to reject. In turn, they have been largely irrelevant for gaining a deeper understanding of the Internet or other large-scale communication networks.

References

[1] D. Alderson, J. C. Doyle, R. Govindan, and W. Willinger. Toward an optimization-driven framework for designing and generating realistic Internet topologies. In *Proc. of ACM SIGCOMM HotNets-I*, 2002.

[2] L. S. Brakmo and L. L. Peterson. TCP Vegas: end to end congestion avoidance on a global Internet. *IEEE Journal on Selected Areas in Communications*, 13(8):1465–80, October 1995. http://cs.princeton.edu/nsg/papers/jsac-vegas.ps.

[3] J. M. Carlson and J. C. Doyle. Complexity and robustness. In *Proc. Natl. Acad. Sci. USA, 99, Suppl. 1*, pages 2538–2545, 2002.

[4] J. M. Carlson and J. C. Doyle. Highly Optimized Tolerance: a mechanism for power laws in designed systems. *Physics Review E*, 60:1412–1428, 1999.

[5] J. M. Carlson and J. C. Doyle. Highly Optimized Tolerance: Robustness and design in complex systems. *Physics Review Letters*, 84(11):2529–2532, 2000.

[6] H. Choe and S. H. Low. Stabilized Vegas. In *Proc. of IEEE Infocom*, April 2003. http://netlab.caltech.edu.

[7] M. E. Crovella and A. Bestavros. Self-similarity in World Wide Web traffic: Evidence and possible causes. *IEEE/ACM Transactions on Networking*, 5(6):835–846, 1997.

[8] J. C. Doyle and J. M. Carlson. Power laws, Highly Optimized Tolerance and generalized source coding. *Physics Review Letters*, 84(24):5656–5659, 2000.

[9] C. V. Hollot, V. Misra, D. Towsley, and W. B. Gong. Analysis and design of controllers for AQM routers supporting TCP flows. *IEEE Transactions on Automatic Control*, 47(6):945–959, 2002.

[10] C. Jin, D. Wei, S. H. Low, G. Buhrmaster, J. Bunn, D. H. Choe, R. L. A. Cottrell, J. C. Doyle, H. Newman, F. Paganini, S. Ravot, and S. Singh. Fast kernel: Background theory and experimental results. In *First International Workshop on Protocols for Fast Long-Distance Networks*, February 2003.

[1] While these rubrics differ from one another, from the perspective of this paper, these differences are all quite minor.

[11] C. Jin, D. X. Wei, and S. H. Low. FAST TCP: motivation, architecture, algorithms, performance. submitted for publication, netlab.caltech.edu/, July 2003.

[12] F. P. Kelly, A. Maulloo, and D. Tan. Rate control for communication networks: Shadow prices, proportional fairness and stability. *Journal of Operations Research Society*, 49(3):237–252, March 1998.

[13] S. Kunniyur and R. Srikant. Designing AVQ parameters for a general topology network. In *Proceedings of the Asian Control Conference*, September 2002.

[14] S. Kunniyur and R. Srikant. A time-scale decomposition approach to adaptive ECN marking. *IEEE Transactions on Automatic Control*, June 2002.

[15] S. Kunniyur and R. Srikant. End-to-end congestion control: utility functions, random losses and ECN marks. *IEEE/ACM Transactions on Networking*, 2003. to appear.

[16] S. Kunniyur and R. Srikant. End–to–end congestion control schemes: utility functions, random losses and ECN marks. In *Proceedings of IEEE Infocom*, March 2000. http://www.ieee-infocom.org/2000/papers/401.ps.

[17] W. E. Leland, M. S. Taqqu, W. Willinger, and D. V. Wilson. On the self-similar nature of Ethernet traffic. *IEEE/ACM Transactions on Networking*, 2(1):1–15, 1994.

[18] L. Li, R. Tanaka, D. Alderson, S. L. Low, W. Willinger, and J. C. Doyle. The importance of annotated graphs in Internet topology generation. submitted for publication, netlab.caltech.edu/, July 2003.

[19] S. H. Low, F. Paganini, J. Wang, and J. C. Doyle. Linear stability of TCP/RED and a scalable control. *Computer Networks Journal, to appear,*, 2003. http://netlab.caltech.edu.

[20] S. H. Low. A duality model of TCP and queue management algorithms. *IEEE/ACM Trans. on Networking*, 11(4):525–536, August 2003. http://netlab.caltech.edu.

[21] S. H. Low and D. E. Lapsley. Optimization flow control, I: basic algorithm and convergence. *IEEE/ACM Transactions on Networking*, 7(6):861–874, December 1999. http://netlab.caltech.edu.

[22] S. H. Low, L. L. Peterson, and L. Wang. Understanding Vegas: a duality model. *J. of ACM*, 49(2):207–235, March 2002. http://netlab.caltech.edu.

[23] L. Massoulie and J. Roberts. Bandwidth sharing: objectives and algorithms. In *Infocom'99*, March 1999. http://www.dmi.ens.fr/%7Emistral/tcpworkshop.html.

[24] L. Massoulie and J. Roberts. Bandwidth sharing: objectives and algorithms. *IEEE/ACM Transactions on Networking*, 10(3):320–328, June 2002.

[25] J. Mo and J. Walrand. Fair end-to-end window-based congestion control. *IEEE/ACM Transactions on Networking*, 8(5):556–567, October 2000.

[26] F. Paganini, J. C. Doyle, and S. H. Low. Scalable laws for stable network congestion control. In *Proceedings of Conference on Decision and Control*, December 2001. http://www.ee.ucla.edu/ paganini.

[27] F. Paganini, Z. Wang, J. C. Doyle, and S. H. Low. Congestion control for high performance, stability and fairness in general networks. submitted for publication, March 2003.

[28] F. Paganini, Z. Wang, S. H. Low, and J. C. Doyle. A new TCP/AQM for stability and performance in fast networks. In *Proc. of IEEE Infocom*, April 2003. http://www.ee.ucla.edu/ paganini.

[29] P. A. Parrilo and B. Sturmfels. Minimizing polynomials functions. Submitted to the DIMACS volume of the Workshop on Algorithmic and Quantitative Aspects of Real Algebraic Geometry in Mathematics and Computer Science. http://xyz.lanl.gov/abs/math.OC/0103170, 2001.

[30] P. A. Parrilo. *Structured semidefinite programs and semialgebraic geometry methods in robustness and optimization*. PhD thesis, California Institute of Technology, http://www.control.ethz.ch/ parrilo/pubs/index.html, 2000.

[31] V. Paxson and S. Floyd. Wide-area traffic: the failure of poisson modeling. *IEEE/ACM Transactions on Networking*, 3(3):226–244, 1995.

[32] G. Vinnicombe. On the stability of networks operating TCP-like congestion control. In *Proc. of IFAC*, 2002.

[33] G. Vinnicombe. Robust congestion control for the Internet. submitted for publication, 2002.

[34] J. Wang, L. Li, S. H. Low, and J. C. Doyle. Can TCP and shortest-path routing maximize utility? In *Proc. of IEEE Infocom*, April 2003. http://netlab.caltech.edu.

[35] W. Willinger and V. Paxson. When mathematics meets the Internet. *Notices of the AMS*, 45(8):961–970, 1998.

[36] W. Willinger, M. S. Taqqu, R. Sherman, and D. V. Wilson. Self-similarity through high variability: statistical analysis of Ethernet LAN traffic at the source level. *IEEE/ACM Transactions on Networking*, 5(1):71–86, 1997.

[37] W. Willinger and J. C. Doyle. Robustness and the Internet: Design and Evolution. In *Robust design: A Repertoire of Biological, Ecological, and Engineering Case Studies*, E. Jen, Editor, Oxford University Press (to appear).

[38] H. Yaiche, R. R. Mazumdar, and C. Rosenberg. A game theoretic framework for bandwidth allocation and pricing in broadband networks. *IEEE/ACM Transactions on Networking*, 8(5), October 2000.

[39] X. Zhu, J. Yu, and J. C. Doyle. Heavy tails, generalized coding, and optimal web layout. In *Proceedings of IEEE Infocom*, April 2001.