Infinitesimal Perturbation Analysis for Stochastic Flow Models: A Tool for Management, Control, and Optimization of Telecommunications Networks

Yorai Wardi
School of Electrical and Computer Engineering
Georgia Institute of Technology
Atlanta, GA 30332
USA
ywardi@ee.gatech.edu

Infinitesimal Perturbation Analysis (IPA) was developed in the nineteen-eighties as a sensitivity analysis method for computing the derivatives (gradients) of sample performance functions, defined on queueing networks, with respect to traffic and structural parameters. Owing to its simplicity and low computational requirements, IPA initially appeared to have the potential of underlying a simulation-based optimization framework for design and provisioning of communications networks. However, it was later discovered that the IPA gradients often were statistically biased in a large class of networks and performance functions, including networks with multiple customer classes, packet loss, and non-Markovian routing. Although various techniques were developed to overcome this problem, they often involve higher computing costs that the basic IPA, or require a knowledge of the underlying probability law underlying the traffic processes.

In recent years, an alternative framework to modeling and simulation of high-speed networks has emerged, based on continuous flow instead of "discrete" packets. Dubbed *Stochastic Flow Models (SFM)*, the framework constructs models which forego the detailed dynamics of each individual packet, and instead focus on the entire flow. From the standpoint of IPA, such models overcome the above-mentioned difficulties in that they provide statistically-unbiased gradients in a far-larger class of networks then packet-based models. Furthermore, they have the following two additional appealing properties: (i) they require low computational resources, and (ii) they are nonparametric and model-free, in the sense that they require no knowledge of the underlying probability law or of a detailed model for the traffic processes. For these reasons they hold promise of real-time implementation in the nodes of a network and of facilitating an approach to optimization-based management and control.

This talk will present the principles of IPA and its application to SFM networks, highlight its salient features via examples, and discuss its potential deployment in real-time optimization of telecommunications networks.