Perspectives on Traffic Modeling in Networks

Peter W. Glynn

Stanford University

2012 Stochastic Networks Conference, June 21, 2012

25 years of stochastic network conferences

since Tom Kurtz and I organized the first one in Madison, WI

- Madison, June, 1987
- Minneapolis, March 1994
- Edinburgh, August 1995
- Madison, June 2000
- Stanford, June 2002
- Montreal, July 2004
- Urbana-Champaign, June 2006
- Paris, Ecole Normale Superieure, June 2008
- Cambridge, Newton Institute, March 2010
- MIT, June 2012

These conferences have significantly impacted the field:

Queueing theory in 1987:

- Early days of heavy-traffic theory, loss networks, etc
- Much work focused on very detailed "closed form" calculation for various variants of the single-server queue
- Transforms, recursive methods, etc were key tools

Today:

- Appreciation of multi-scale phenomena is central
- Stochastic geometry, random graphs, etc play a key role in many models
- Rich connections to probability, theoretical computer science, etc
- Modeling tends to be more ambitious $\rightarrow$ bigger questions to be answered
- Algorithm design as an endpoint; modeling as a vehicle to design algorithms
- More focus on qualitative insights; use of asymptotic regimes to study optimality issues

What has made these conferences successful?

- Brings together a diverse and active collection of researchers, cutting across multiple communities having an interest in such models
- Has been highly responsive to the changing applications landscape
  - Circuit-switched networks
  - ATM networks
  - Wafer fabs
  - Internet/TCP
  - Call centers
  - Mobile networks
  - Sensor networks
  - Peer-to-peer networks
  - Data centers

Many important ideas and methods have been promoted at this conference

- Multi-class stability issues
- Connecting stability to fluid models and LPs
- Understanding heavy traffic in single-server and many-server settings
- State space collapse/snapshot principle
- Use of economic principles to guide the construction of distributed algorithms
- Use of randomized algorithms to lower computational/communications costs
- Large deviations, random graphs, etc

The rest of this talk...

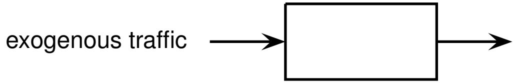
Perspectives on traffic modeling:

- A view or vista
- A mental view or outlook
- The relationship of aspects of a subject to each other and to a whole
- Subjective evaluation of relative significance; a point of view

Something fun to discuss over lunch today...

What is a traffic model?

- The exogenous input to a network



Three main uses of models:

- Descriptive
- Predictive
- Prescriptive

The choice of traffic model can impact conclusions in all 3 settings...

The Simplest Traffic Model: Poisson Arrivals

Justified by "Superposition Theorem"

- Large number n of independent sources, none of which contributes a significant percentage of the overall traffic
- Describes traffic at the time scale of individual interarrival times
- Does NOT describe traffic at larger time scales ("Poisson breakdown phenomenon")

Analog to superposition theorem establishes that for many-server queues in heavy-traffic, abandonment process is "doubly stochastic" Poisson

The Renewal Model:

- Other than Poisson case, difficult to conceive of a mechanistic explanation for renewal arrival epochs

Correlated Arrival Stream:

- Often realistic as a means of incorporating "burstiness" in traffic
- Markov-modulated Poisson processes
- Long-range dependence

To the degree that we build "Markov-modulated burstiness" into our models, the "standard asymptotics" suppress the presence of the burstiness:

- If $N(\cdot)$ is the associated counting process,

$$N(t) \approx \lambda t + \eta\sqrt{t}N(0,1)$$

- Fluid limits, diffusion limits, etc look qualitatively identical to what is seen in the renewal case

- Describes a very "statistically regular" world

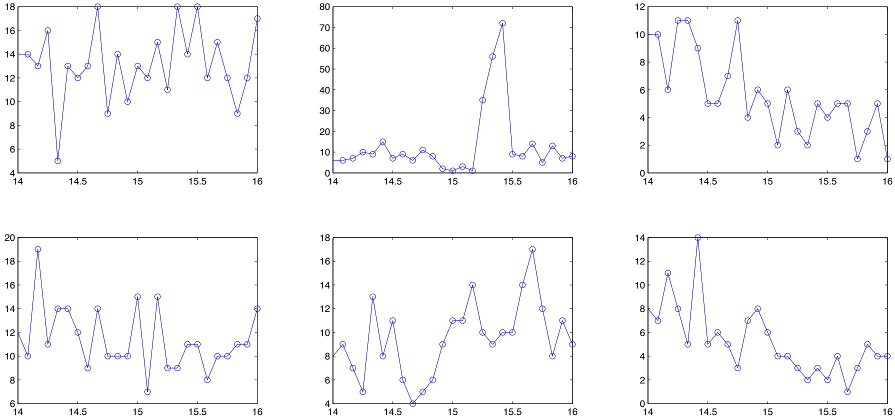
- Explanation: The "decorrelation time" for the underlying Markov modulation is $O(1)$; the Markov modulation describes bursts that are $O(1)$

Now, feed the traffic into a network in "heavy traffic": "Relaxation time/time to equilibrium" of order $1/(1-\rho)^2$ but bursts are $O(1)$ in duration

This asymptotic is relevant to a model with (very) short-lived bursts; same applies to large deviation paths to overflow in large buffer regime

But there are problem settings where burstiness persists over long time scales...

Six Different Tuesdays



In call center setting:

- Relaxation time/time to equilibrium for many-server systems is $O(1)$
- But bursts persist over time scales that are $O(1)$
- And time-of-day effects kick in over time scales that are $O(1)$

A Statistical Analysis (joint work with J. Hong, X. Zhang):

- We want a process that is "locally Poisson"
- Our choice:

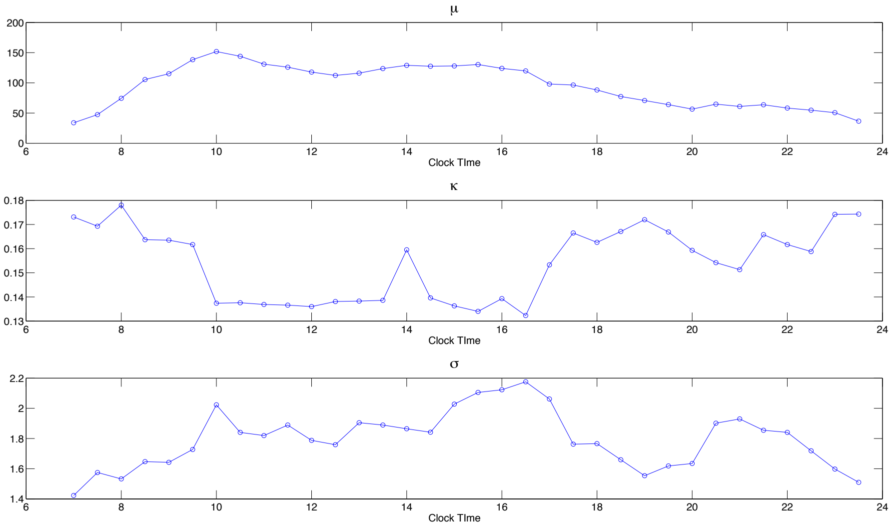
$$N(t) = \widetilde{N}\left(\int_0^t X(s)ds\right)$$

where $\widetilde{N}$ is a unit rate Poisson process independent of X, where X satisfies

$$dX(t) = \kappa(t)(\mu(t) - X(t))dt + \sigma(t)\sqrt{X(t)}dB(t)$$

with $\kappa(\cdot)$, $\mu(\cdot)$, and $\sigma(\cdot)$ piecewise constant

- Low parameter model
- Exhibits mean reversion
- Analytically tractable (affine process)



Mean reversion time is of order over which time-of-day effects manifest (fit by "generalized method of moments")

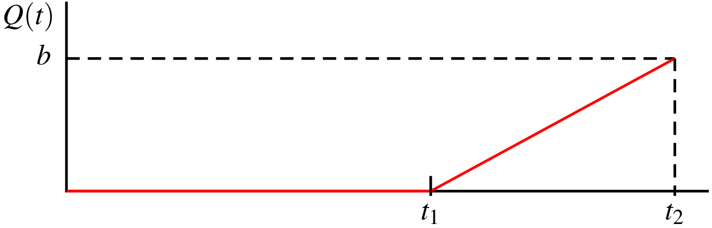
At some level, the "dominant randomness" in this model comes from unpredictable bursts that are not captured in "standard asymptotics"

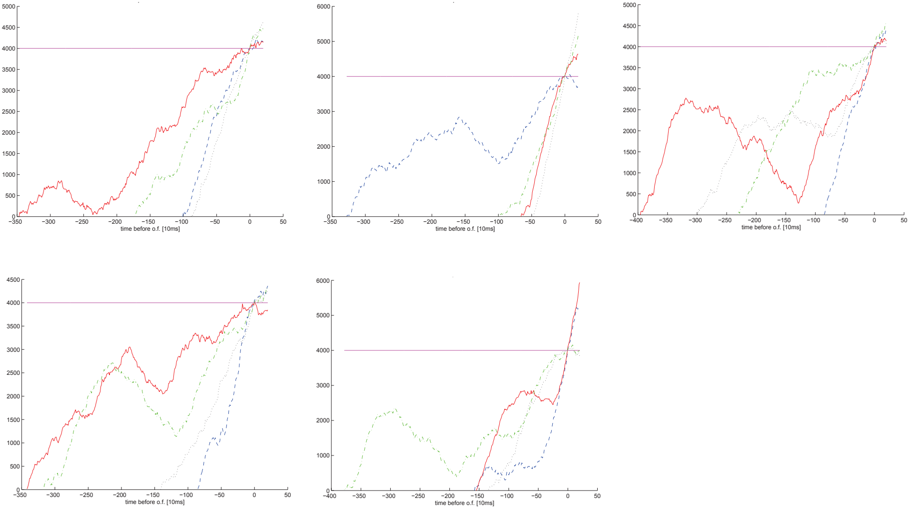
Question:

How does presence of long time-scale burstiness impact the qualitative conclusions we reach from traditional models?
(e.g. "square-root staffing")

Another example to illustrate this point...

Application of large deviations to traditional traffic models predicts very specific "conditional dynamics" for rare events (e.g. buffer overflow)





- Traffic source: UNC Network Data Analysis Study Group (2003).
- Buffer size = 4,000 packets (about 60 ms worth of traffic).

These paths are not consistent with the "rare event behavior" of any of the known traffic models

Question:

- Should we worry about this "disconnect" between the qualitative behavior predicted by our theory or not?
- If not, why not? If yes, what should we do about this?

As the above illustrates, our area has plenty more to do...

- There will be new technologies and new settings in which networks appear
- And stochastic features will often be a first-order effect that needs to be modeled

So, I expect many more excellent SNC's in the years to come!