



# Modeling for Performance Analysis, Network Engineering, and Cybersecurity



Ian Akyildiz, Georgia Tech  
John Baras, University of Maryland  
William Cleveland, Purdue University  
Jennifer Neville, Purdue University  
Ness Shroff, Ohio State University  
Jian Tan, Ohio State University  
Mark Ward, Purdue University



# Three Types of Modeling

---

Technology-based specifications, deterministic and probabilistic, of network protocols, device architectures, and traffic

- based on knowledge of networking
- bring as much generality as possible to probabilistic modeling to make applicability as wide as possible
- generality sometimes creates much more challenging math problems

Empirical modeling based on data, much in the form of packet traces

- build statistical (probabilistic) models
- estimate parameters of the models and characterize statistical variability of estimates
- background of guidance from networking

Hybrid

- combination of technology and empirical modeling

# Modeling Research Projects



## Traffic Modeling and Performance Analysis for Multiscale Networks (Technology)

- John S. Baras ISR, ECE, BiE Depts and AMSC, University of Maryland
- mobile stations use the same wireless channel to send packets to the same destination
- model probability that at each time slot a station has a packet to send
- if channel is busy must wait a back-off time  $t_u$  with a probability distribution modeled in various ways
- channel service time  $t_s$  depends on a packet size with a probability distribution modeled in various ways
- study distribution of packet transit time to destination

## Retransmissions, Heavy Tails, and Long-Range Dependence (Technology)

- Jian Tan and Ness Shroff, ECE & CSE, Ohio State
- channel with on-off intervals modeled by probability distributions
- packet sizes modeled by probability distribution over finite interval
- channel errors modeled to depend on size
- study probability distributions of time until successful packet transmission, and number of transmissions

# Modeling Research Projects



## Spatial Correlation and Mobility Aware Traffic Modeling for Mobile Dense Wireless (Sensor) Networks (Technology)

- Ian F. Akyildiz, BWN Lab, School of ECE
- tactical mobile network with a commander node and nodes communicating with commander
- model movement of nodes by spatially correlated processes with pause periods and movement periods
- traffic for each node an on-off process transferring files of different sizes
- study traffic properties for prediction and control

## Multifractal and Gaussian FSD Models for Best-Effort Traffic (Empirical)

- D. Anderson, Math, Xavier University; B. Xi, W. S. Cleveland, Statistics, Purdue
- models for multiplexed traffic interarrival times that have a very simple structure
- provide an effective time-domain intuition for traffic phenomena
- allow derivation of formulas for traffic phenomena and other mathematical investigations
- use to generate traffic for simulations for network engineering

# Modeling Research Projects



## Statistical Analysis and Modeling of VoIP (SIP-RTP) Traffic (Empirical)

- B. Xi, W. S. Cleveland, Statistics, Purdue; H. Chen, Yahoo; Thomas Telkamp, Cariden Inc.
- models for arrival times of multiplexed VoIP packet traffic
- semi-empirical statistical model and a mathematical statistical model
- first reported comprehensive analysis and modeling of VoIP traffic from a large ISP with all VoIP applications
- use to generate traffic for simulations for network engineering

## VoIP Network Traffic Engineering: Analysis of Queueing Delay and Jitter (Empirical)

- J. Xia, W. Cleveland, and M. Ward, Statistics, Purdue
- use semi-empirical model to generate arrival times at different traffic rates
- queueing simulation for utilizations 0.1(0.1)0.9 for each rate
- analyze entire distribution of queueing delay and jitter
- seek statistical models and insight for distributions
- use mathematical results on queueing for guidance (e.g., Choe and Shroff, 1999)
- determine utilizations by rate that satisfy delay and jitter requirements
- Phase I of long-term project

# Modeling Internet Connection Traffic for Cybersecurity: The Traffic



Connection: defined by 5-tuple (source/destination IP-address/port-number and protocol)

Unidirectional or bidirectional

Associated with each connection are variables that depend on the security task

Examples of variables

- the 5-tuple
- the time of the first packet
- total bytes in each direction
- total packets in each direction
- application (e.g., http)
- connection initiator
- round-trip time for TCP
- number of ssh client keystroke packets of any connection from a packet-level keystroke algorithm

Example database: all connections using the gateway between an “inside” network and the outside

# Modeling Internet Connection Traffic for Cybersecurity: Data Structures



## A marked point process

- arrivals are timestamps of connections
- multivariate marks are the variables: 5-tuple and other variables

## A marked graph

- nodes are the unique IP addresses of the connections
- edges between two IP addresses that are source and destination of at least one connection
- marks at an edge are values of connection variables for connections involving the two IP addresses

# Modeling Internet Connection Traffic for Cybersecurity: Security Tasks



Detect login intrusions

Early detection of attacks producing high volume of packets or application connections

Automated defensive actions

Rapid human forensics

Source-destination surveillance

# Modeling Internet Connection Traffic for Cybersecurity: Status



Modeling connection traffic for cybersecurity is just beginning.

# Modeling Internet Connection Traffic for Cybersecurity: Status

---



## Cybersecurity Plan to Involve NSA, Telecoms DHS Officials Debating The Privacy Implications

By Ellen Nakashima  
Washington Post Staff Writer  
Friday, July 3, 2009

The Obama administration will proceed with a Bush-era plan to use National Security Agency assistance in screening government computer traffic on private-sector networks, with AT&T as the likely test site, according to three current and former government officials.

President Obama said in May that government efforts to protect computer systems from attack would not involve "monitoring private-sector networks or Internet traffic," and Department of Homeland Security officials say the new program will scrutinize only data going to or from government systems.

# Modeling Research Projects



## A Rules-Based Statistical Algorithm for Keystroke Detection (Hybrid)

- P. Kidwell, S. Guha, W. Cleveland, Statistics, Purdue; J. Gerth, CS, Stanford
- identifies ssh client keystroke packets in any TCP connection
- uses packet sizes, direction, flags, and interarrival times
- connection with  $\geq 4$  keystrokes classified as interactive login
- applied to all connections (not just port 22) seen by a network monitor at the gateway of an “inside” network
- excellent performance compared with past methods

## Relational Poisson Models for Anomaly Detection (Empirical)

- Jennifer Neville and Nesreen Ahmed, CS, Purdue
- database: inside-outside database
- model counts with non-homogeneous, time-modulated Poisson distributions
- exploit relational dependencies by model dependencies of the parameters across pairs with a common destination IP address
- characterize count statistical properties to detect anomalies