CSCE 619-600
Networks and Distributed Processing
Spring 2017

Introduction

Dmitri Loguinov
Texas A&M University

January 17, 2017

Agenda

- Course overview
- Homework requirements
- What is modeling?
- What are simulations?
- Examples
- Wrap-up

Course Overview

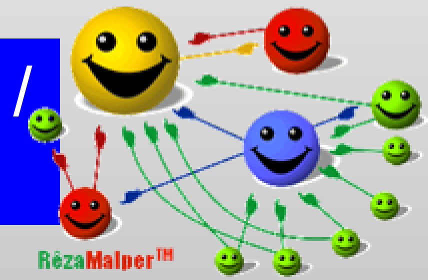
- This course focuses on analytical modeling and simulation of computer systems/networks
- Since many events computer science are random, a huge part of modeling relies on various types of probability theory
  - Also true for other engineering fields
- There will be a review of basic probability theory, but it helps if you remember some of it
  - Other pre-requisites: undergraduate calculus, some matrix algebra, basic graph theory, programming

Course Overview 2

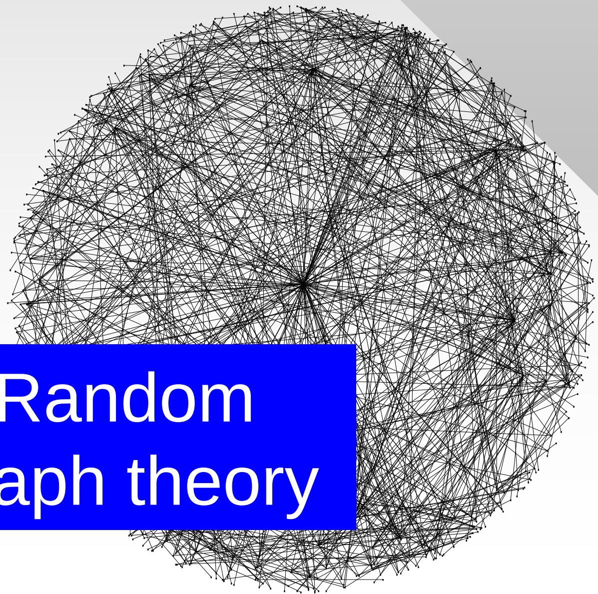
- This class is a mix of several applied topics
  - Probability theory (review), stochastic processes, Markov chains, random graph theory, and control theory

Review of probability theory

Renewal theory / Markov chains



midterm



Random graph theory

Congestion control

final

Course Overview 3

- Some of the homework problems
  - Bus wait: you randomly walk to a bus stop where the average delay between buses is 20 min; how long is your wait?
  - Bank robbery: suppose it takes 6 minutes to rob a bank; police periodically drive by (mean delay 20 min); what is the probability the robber is caught?
  - TV surfing: you randomly flip channels and stumble onto a movie whose duration is X; you watch it until it ends or you get bored after T time units; both X,T are random variables; how long will you be watching?
  - Save the forest: fire can jump between trees if they are within 20 feet of each other; given an area of 100x100 miles with K trees, what is the probability that one randomly ignited tree burns down the whole forest?

Syllabus

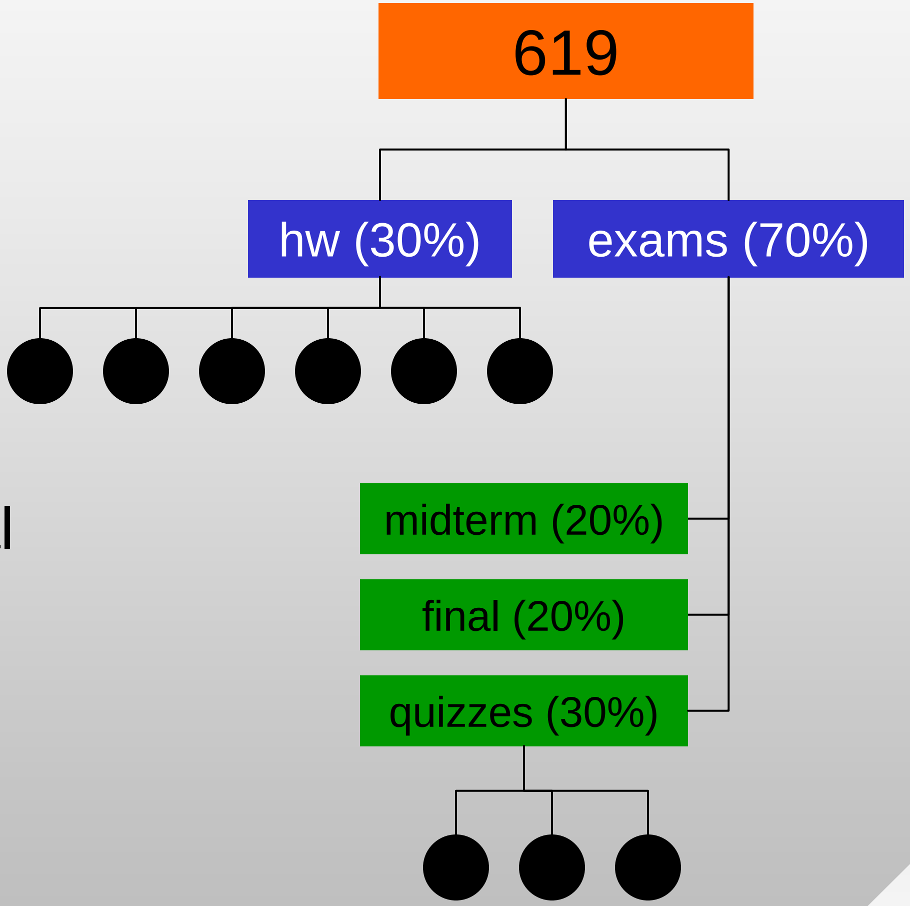
- Homework is a combination of simulations and analytical derivations
  - All homework must be accompanied by a clearly written report in Latex that explains your results and simulation setup
- Team work is not allowed
- Reminder: you may not pass any material from the web, other students, or publications as your own
  - Penalty for cheating is an F*
  - See rule 20 at http://student-rules.tamu.edu/

Syllabus 2

- Office hours
  - TR 5:10-6:10pm in HRBB 515C
  - Website: http://irl.cse.tamu.edu/courses/619
  - Forum: http://piazza.com/tamu/spring2017/csce619
- All lectures and homework on the website
  - Including hints on using Latex and various support files
- Final grades
  - A: 80-100%
  - B: 70-79%
  - C: 60-69%
  - D: 50-59%
  - F: 0-49%

Syllabus 3

- Assignments/exams
  - Midterm: 20%
  - Final: 20%
  - Quizzes (3): 30% total
  - Homework (6): 30% total
- Quizzes cover
  - Probability theory
  - Renewal processes
  - Random graphs
- Exams cover half a semester each



Syllabus 4

- Recommended reading:
  - R.W. Wolff. *Stochastic Modeling and the Theory of Queues.* Prentice Hall, 1989
  - B. Bollobas. *Random Graphs*. Academic Press, 2001
  - R. Srikant. *The Mathematics of Internet Congestion Control.* Birkhauser, 2004
  - J.F. Kurose and K.W. Ross. *Computer Networking: A Top Down Approach Featuring the Internet*. Addison-Wesley, 6th edition, 2012 (general undergraduate background)

What is Modeling

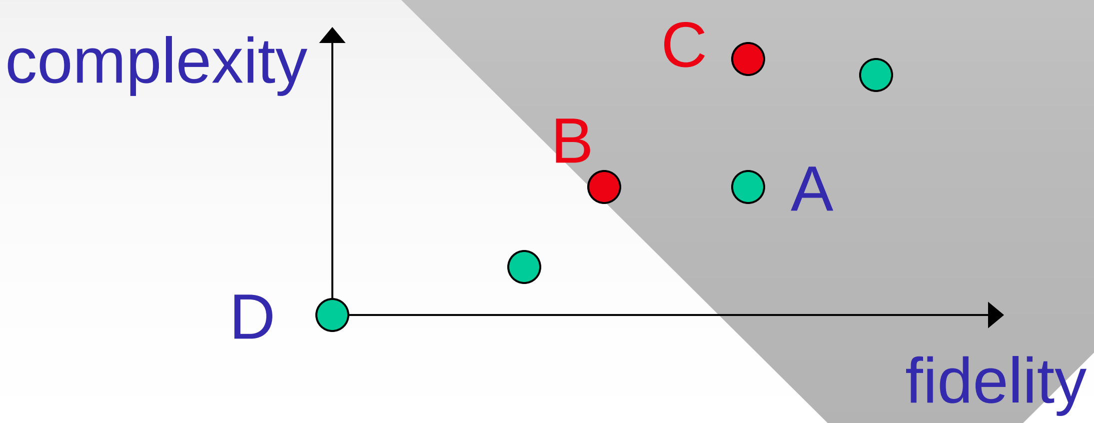
- Example: can we determine if a particular person will have a car accident today (e.g., for insurance purposes)?
  - Depends on the time they leave for work, route taken, speed at different times t, number of cars encountered, delay at each intersection, decision-making, etc.
  - This deterministic system is very complex
- In research, we aim to understand the behavior of complex systems and then hopefully improve them
  - To do this, we first need to describe the system in mathematical terms, or create a *model* for it
  - Most models are approximations to real behavior

What is Modeling 2

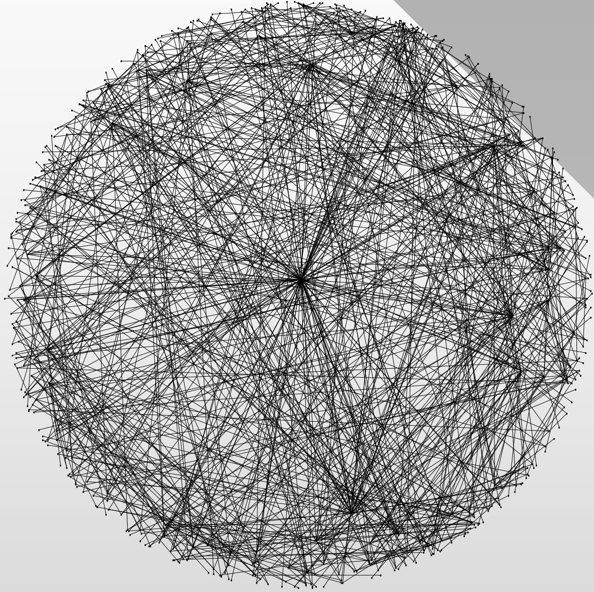
- Complex deterministic systems often replaced with much simpler stochastic ones
  - Back to our example: insurance companies assign accident probabilities to different drivers
- Even simple systems require a model, but we sometimes neglect this
- Example: you have x apples and you sell half of them
  - How many apples are you left with?
  - What assumptions did you make?

What is Modeling 3

- Building models
  - Involves a tradeoff between complexity and fidelity
- Complexity means how difficult it is to obtain the parameter in question from the model
- Fidelity is how far this parameter deviates from that in real systems and under what assumptions



Simulations



↑

simulated Internet

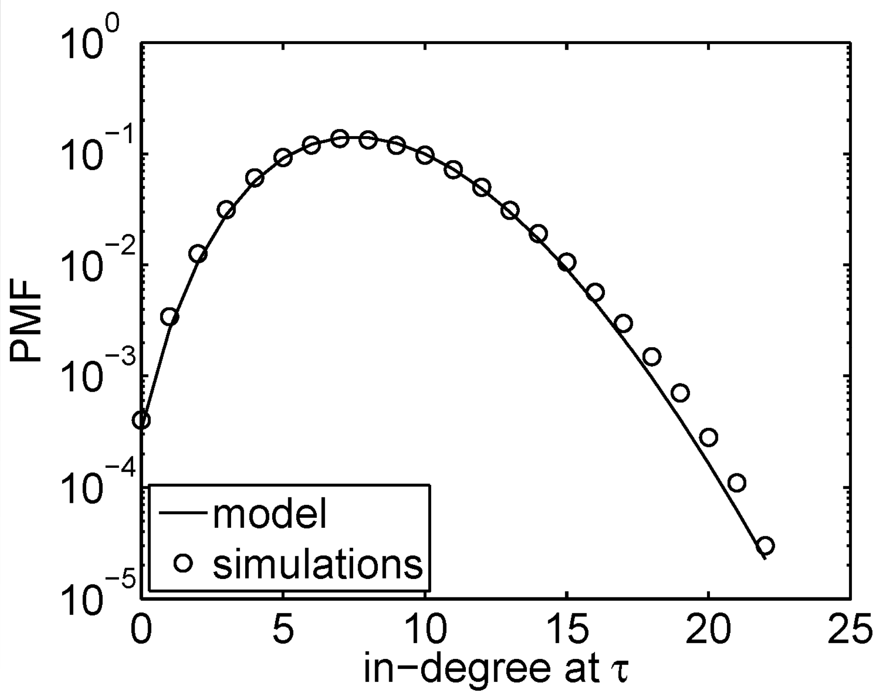
- Models need to be verified
  - Either in real systems or simulations
- Simulations are easier
  - Real-life experiments are usually costly, hard to control, and generally non-repeatable
- What is a simulation?
  - Execution of an algorithm to produce an estimate of parameters in question
- Use a good number generator

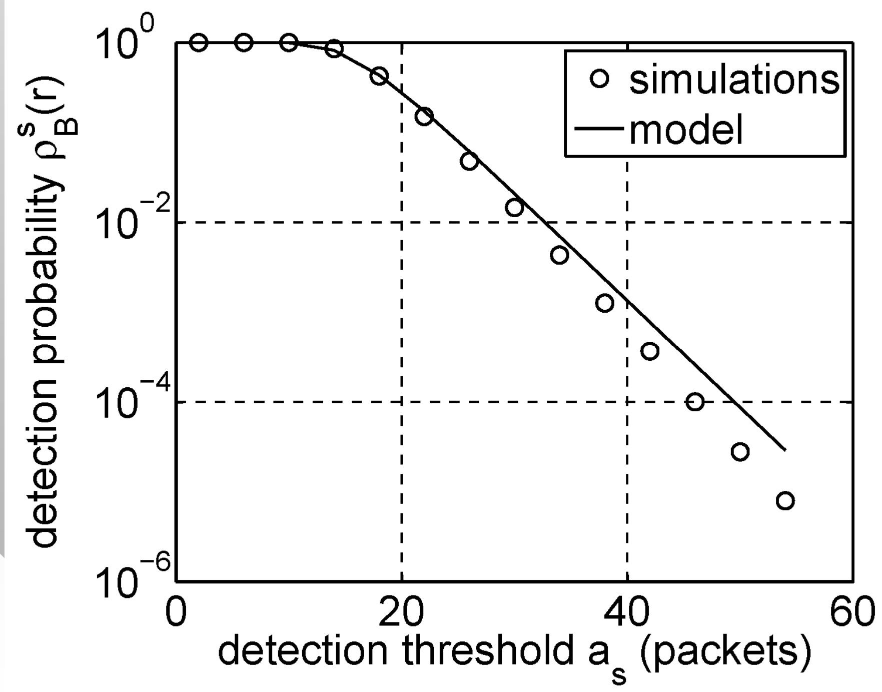
```
init_genrand ((DWORD)time(NULL));
double u = genrand_res53 ();
```

  - See course website (Mersenne Twister for C/C++)

Simulations 2

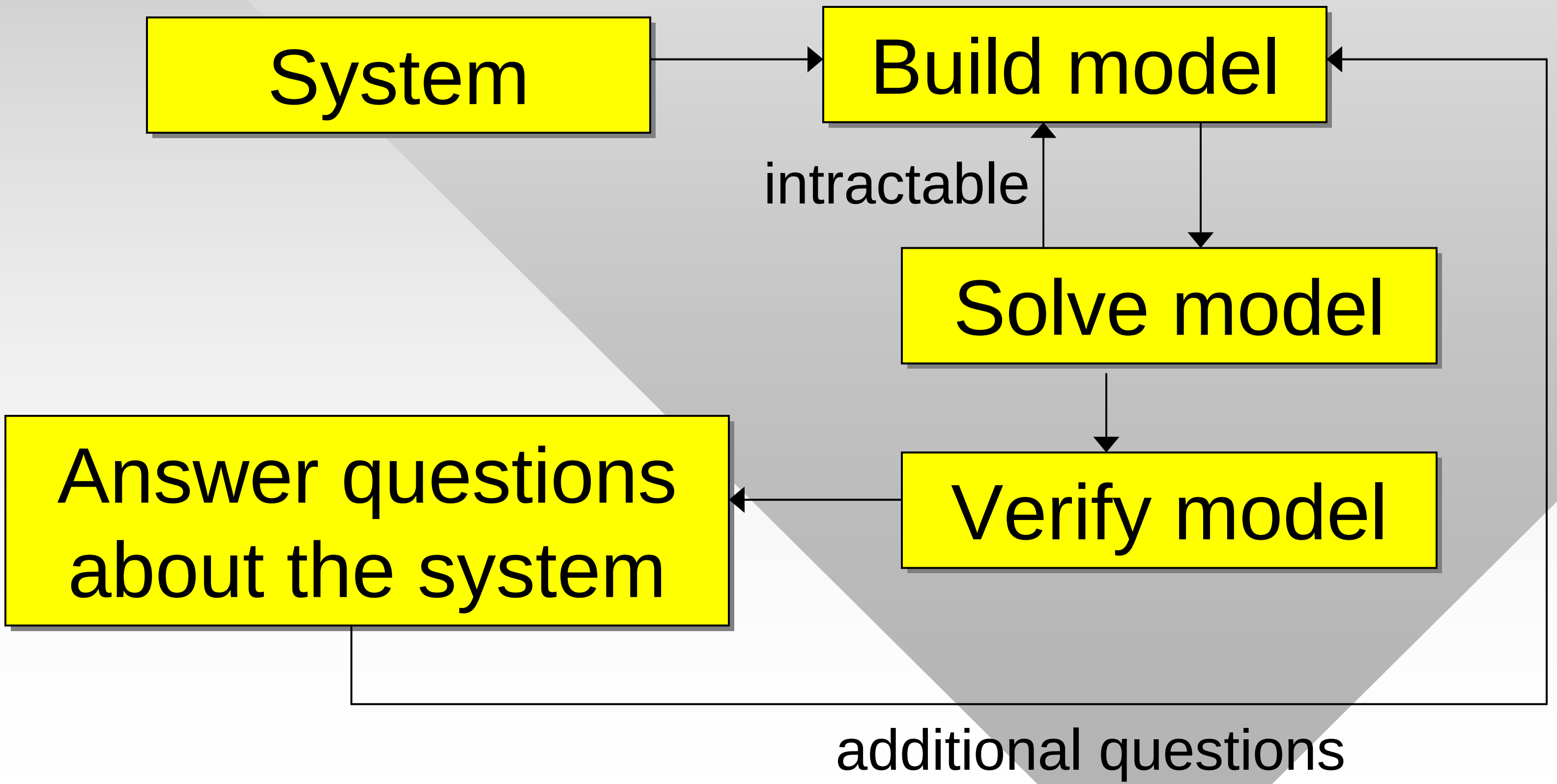
- Plotting results
  - When comparing results against a model, follow the examples below (circles for discrete points, solid curves for continuous functions)





Summary

- Modeling is a reduction of your system to some set of equations that allow one to obtain knowledge about its behavior



Renewal Example

- Bus-wait problem
  - Imagine a bus stop and a sequence of buses arriving to it with inter-bus delays X_1, X_2, …

bus $i+1$ | bus i | bus $i - 1$

- Distribution of X_i is unknown and only its average $E[X_i] = s = 20$ minutes is posted
  - Q1: Determine your expected wait time from a random point t when you approach the bus stop

Renewal Example 2

- Common sense suggests half of $E[X_i]$, i.e., 10 min
  - Can we prove this rigorously?
  - How accurate is this model?
- This is true if buses arrive exactly every 20 minutes, but does not hold in any other case
- Inter-bus delays are uniform in $[0, 2s]$
  - What is the expected wait time?

$$P(X_i < x) = \frac{x}{2s}$$

- What if they are exponentially distributed?
  - Is the wait time more or less on average in this case compared to the previous two?

$$P(X_i < x) = 1 - e^{-x/s}$$

Renewal Example 3

- Suppose that inter-bus delays are Pareto (from a class of *heavy-tailed* distributions):

$$P(X_i < x) = 1 - (1 + x/\beta)^{-\alpha}$$

  - where α is the shape parameter, β is the scale parameter, and the mean inter-bus delay is $E[X_i] = \beta/(\alpha - 1)$

- Interestingly, under Pareto inter-bus delay, your expected wait time is more than $E[X_i]$
  - The heavier the tail, the longer the wait
  - For $\alpha = 3$, your expected wait is $2s = 40$ minutes
  - For $\alpha = 2.5$, it is $4s = 80$ minutes, and so on
  - For $1 < \alpha \leq 2$, the average wait time is infinity

Renewal Example 4

- Another question is related to conditional expectations
  - Q2: Suppose you know that the last bus left 5 minutes ago, what is your expected wait delay now?
- Constant X_i is easy
  - Simply 15 minutes




- Uniform X_i
  - Wait time uniform in $[0, 35]$
  - Expected wait time is $E[X_i - 5 \mid X_i > 5] = 17.5$ minutes




- Exponential X_i
  - Still 20 minutes

Renewal Example 5

- For Pareto X_i and $\alpha = 3$, expected wait is 22.5 min
- Well, now suppose the last bus left $t = 1$ hour ago
  - What is the wait now? Exponential? Pareto?
- In the $\alpha = 3$ Pareto case, the wait now is 50 min
  - For $t = 2$ hours, your expected wait is 80 minutes
  - The longer you've waited, the longer you will continue waiting on average (inspection paradox)

- Intuition
  - You're likely to miss sequences of buses with very small delays and arrive during a very long inter-bus interval

Renewal Example 6

- Solutions to these problems are studied in renewal process theory and Markov chains, which can be used to model a variety of systems
  - Packet arrivals and queuing delays in routers (queuing theory, Markov chains)
  - User and job arrivals into computers, web servers, distributed systems, social and peer-to-peer networks
  - Google PageRank model of users randomly browsing the web, which gives higher weight to important nodes
  - Many other recurring phenomena
- In the homework, we use renewal process theory to study resilience of P2P systems to disconnection