Analysis and Prediction of the Dynamic Behavior of Applications, Hosts, and Networks

Syllabus

Web Page

http://www.cs.northwestern.edu/~pdinda/predclass-w02

Instructor

Peter A. Dinda
1890 Maple Avenue, Room 326
847-467-7859
pdinda@cs.northwestern.edu
Office hours: 2-4pm on Wednesdays or by appointment

Location and Time

1890 Maple Avenue, CS Department small classroom (342), WF 10:30-12

Prerequisites

Required	CS 311 or equivalent data structures course
Required	Calculus and some linear algebra
Recommended	Basic Statistics and Probability
Recommended	Familiarity with Matlab, Maple, and S-Plus

Permissions

CS Graduate Students	No special permissions required (see Pam)
Others	Permission required (see me)

Objectives, framework, and caveats

Important, measurable properties of distributed computing environments, such as application workloads, network bandwidth, and host load, vary dynamically over time. These properties drastically affect the performance of applications running on these systems. However, if applications are made aware of this dynamicity, ideally in the form of statistically meaningful predictions, they often can adapt their behavior to nonetheless provide consistent high performance. In effect, such predictions enable them to exploit the degrees of freedom available in a distributed computing environment to ameliorate the dynamicity found there. In addition, the analyses upon which prediction is based can tell us valuable and fascinating things about the nature of computer systems and networks qua natural systems.

This course focuses on how we can measure, analyze and predict the dynamic behavior of distributed computing environments and their applications. For the most part, we will use an approach based on probability, statistics, and signal processing, although we will also touch on queuing theory and other approaches. The course has three objectives. The first objective is for you to learn some of the theory behind measurement, analysis, and prediction. The second objective is for you to learn how this theory has been applied to computer systems in the past, and what fascinating new things were learned. The final objective is for you to become comfortable in applying the theory to his or her own data and systems, and in evaluating other methods for studying your data.

We will generally read about 2 papers or equivalent materials for each session, covering fundamental ideas and important recent results. Each paper will be presented to the group by a student and then discussed in a round-table manner. In parallel with the readings, students will be strongly encouraged to apply what they are learning by using analytical tools such as Matlab, Maple, Prophet, and others to study real data, ideally data that they themselves are interested in. Students will also be encouraged to play with on-line measurement and prediction systems such as RPS, Remos, and NWS. Finally, each student will complete a quarter-long project in which they will apply what they learn to an area that interests them. The goal of these investigations will be to produce interesting new research results, perhaps even some that will lead to publications.

This is a graduate course and all students in it will be treated like graduate students. I will assume that you are interested in this material, that you can motivate yourself to learn about it, and that you will not be afraid to venture into uncharted territory (i.e., do research). The undergraduate section will differ primarily in that the expectations for the project and for classroom interaction will be slightly lower.

Rumors that we will learn how to predict the stock market have been greatly exaggerated.

Reading

There is no textbook for this course. We will be reading original research papers that report on studies of real computer systems, networks, and their applications. In addition to these papers, there are a number of books, papers, and other resources that are very helpful in understanding the theoretical and statistical analysis techniques that were applied to produce their results. We will read some of this material in tandem with the research papers so that you will be able to generalize the techniques and learn to apply them to your own work. Most of what we will read is available on the web, and I will hand out photocopies of what is not.

Project

Over the course of the quarter, you will apply what you learn to a project of your choice, and then document your project in a high quality paper and presentation. Project topics will be chosen in consultation with me. Projects are to be done individually, although I will discuss exceptions for particularly large or interesting project areas. The expectation for graduate students is that the project will be quality work that the student would not be embarrassed to submit to a workshop. The expectation for undergraduates is that the project be something they would be proud to list on their resumes. All projects will be presented at a public colloquium.

Example project ideas are listed in a separate handout. Because of the high expectations placed on the project, it is vital that you choose to work on something that interests you deeply.

Exams

There will be no exams

Grading

50 % Project
10 % Project paper and presentation
20 % In-class paper presentations of papers
20 % General classroom participation

Schedule

Lecture	Date	Topics	Theory Readings	Application Readings
1	4/3	Mechanics, motivation, overview, calibration, probability/stats review	Jain handouts; Statsoft/Splus textbooks	
2	4/5	Distributions, summaries, estimates, and their implications for the web	Jain handouts; Statsoft/Splus textbooks	Myers (32), Arlitt (35), Smith (40)
3	4/10	Distributions, summaries, estimates, and their implications for hosts		Eager (2); Leland (1), Mutka (3),
4	4/12	Distributions, summaries, estimates, and their implications for hosts **PROJECT PROPOSAL DUE**		Harchol-balter (4), Dinda (5)
5	4/17	Heavy tails, power laws, and self-similarity	Bassingthwaighte, possibly chapter from Mandelbrot	Willinger (11), Paxson (9)

6	4/19	**CANCELLED**		
7	4/24	Self-similarity and power laws in networks		Willinger (12), Faloutsos (17)
8	4/26	Self-similarity in video Nonstationarity **PROJECT UPDATE DUE**		Garrett (30), Cao (24)
9	5/1	Application behavior		Bavier (36), Kapadia (53), Subhlok (58)
10	5/3	Application behavior		Mitra (42), Abdelkhalek (43)
11	5/8	Application behavior		Bhola (45), Keeton (44),
12	5/10	Predicting hosts		Wolski (6), Dinda (7), Devarakonda (48)
13	5/15	Predicting networks		Wolski (15), Basu (13), Sang (23)
14	5/17	Wavelets (Time series of the ‘80s)	Wavelet intro	Ribeiro (19), Feldmann (77)
15	5/22	Chaotic dynamics	Abarbanel (74)	Abarbanel (74), Hofmeyer (57)
16	5/24	Systems **PROJECT UPDATE DUE**		Lowekamp (59), Stemm (63)
17	5/29	Systems		Wolski (62), Dinda (61)
18	5/31	Network paths and links		Downey (26), Allman (27)
19	6/5	Wireless networks		Balachandran (25), Eckhardt (76)
20	6/7	Zipf, file system prediction		Breslau (33), Kroeger (50), Douver (54)
Spares		Anomaly detection		Dasgupta (75), Barford (28)
	6/10-6/14	**Project Papers Due, Project Presentations (TBA)**		