

# **Analyzing the Role of Measurement and Modeling of Internet Traffic in Network Experimentation**

Students: **Rebecca Crabb, Wai Yau, and Stephanie Zolayvar**

Faculty Advisor (PI): **Dr. Kevin Jeffay**, Gillian T. Cell Distinguished Professor of Computer Science, University of North Carolina at Chapel Hill (UNC-CH)

Co-PI: **Dr. Jay Aikat**, Research Assistant Professor, Department of Computer Science, UNC-CH

## **Problem Description**

This CREU project is designed as part of a larger, ongoing research and pedagogical undertaking by the networking research group at the Department of Computer Science at UNC – Chapel Hill (UNC-CH). In this section, we briefly describe this larger project.

Networking researchers and practitioners have been using simulators, experimental research testbeds, and distributed systems for designing and evaluating new networking technologies for more than a decade now. Indeed, experimentation, either via software simulation using simulators such as the Network Simulator (NS) [NS], or via hardware emulation using laboratory testbeds, has been the primary means for evaluating existing and newly proposed protocols and algorithms for improving the Internet. Nonetheless, experimental networking remains a challenging endeavor. Floyd and Paxson cataloged some of the fundamental issues in experimental networking in their seminal paper “Difficulties in Simulating the Internet” [FP01].

Since that time, numerous research projects funded by the NSF, including the Global Environment for Network Innovations (GENI) projects and the traffic generation project at UNC have sought to address some of the difficult methodological issues raised by Floyd & Paxson. However, despite these efforts, the networking research community lacks a coherent, shared view of best practices for experimental methods for networking research. And the teaching of such experimental methods lags even further behind. Even today, there are no agreed-upon research methods or standard practices for maintaining traffic datasets, building networking research testbeds, generating traffic, emulating network characteristics, or running experiments. Thus, while many researchers propose new and improved protocols for improving our cyber-infrastructure, the lack of standards for experimental methods makes real progress and deployment of the best protocols either slow or non-existent.

Thus, the PIs have undertaken a larger and longer-term project of advancing the standards for better research methods in network experimentation. The CREU project envisioned in this proposal is part of that larger project.

## **The CREU project description**

The CREU project is specifically designed to be self-contained and targeted to providing a positive, rich, and realistic research experience for three undergraduate students. The CREU

project has two major components: the first component, designed for completion in late Fall 2011, will be completely collaborative and provide the background learning and information that will equip the students to launch their individual research projects.

The second component consists of specific research questions to be answered by each student, and will involve one of two major research paths: (i) designing and running experiments on research testbeds, and analyzing the results, or (ii) collecting Internet traffic data and conducting a statistical analysis of the data. Each student would have the option of choosing one of these research paths and a specific project in that path. Each student would conduct the research, write a report, and then present her research individually, as part of the departmental requirements for her undergraduate honors thesis in late Spring.

The two major components mentioned above – survey paper in Fall and a hands-on research project in Spring – are described in detail in the section below. Both components are designed such that they will potentially result in research papers. These projects would normally require a huge amount of funding for infrastructure and a lot of time to setup and yield results. However, since the faculty advisors have both invested years of research in this area, the undergraduate students will be able to leverage the existing laboratory resources to hit the ground running.

### **Specific Questions / Hypotheses To Be Investigated**

This project has two main components:

1. Survey of network experimentation: This component will be common to all three students, where they will work collaboratively to learn about the state of the art in network experimentation and document the same in a paper.
2. Honors thesis project: This component is designed so that each of them works on an individual research project that builds upon the previous component, and results in an undergraduate honors thesis for each student. This will also involve setting up a laboratory testbed network, which they will all work together to accomplish. The network setup is intended as a hands-on experience that is necessary for their individual research projects.

This CREU project is designed to serve the following purposes:

- The common component – conducting a survey of network experimentation and writing a paper – will foster collaboration and create a foundation for each student to complete their individual honors thesis. The students will examine the state of the art for experimentation in networking research today. This will involve reading a large set of papers (split among the three students), and then writing up a survey paper that documents networking experimentation as presented in leading conferences – SIGCOMM, IMC, CoNEXT, and the CCR journal during 2009-2011. The students will gain an overall understanding of the state of the art for network experimentation and currently practiced experimental research methods in networking. They will study all aspects of experimentation – network setup/topology, input data for traffic, traffic generation, parameters used, how is the traffic modeled, methods of measurement (and what is measured) for input data as well as experimental output, and the metrics of performance used for assessing outcome.
- Their individual honors projects will give each student a specific research question to explore. This will involve designing experiments, setting up and running experiments, collecting data, analyzing this data, and writing up a report. If the student chooses to do a

statistical analysis of network data, that will involve data collection and significant statistical analyses. Finally, as per departmental requirements for an honors thesis, each student will present her work in an annual undergraduate research symposium held at the end of Spring 2012, which is open to the public.

As examples for inspiration, the PIs will provide possible research ideas to the students for further exploration. Each student, however, will be free to choose from among these ideas or come up with their own research questions. For example, here are four ideas in network experimentation:

1. how does experimental outcome differ when using same input traffic data using non-parametric modeling vs. parametric modeling for generating application data in traffic generation? Leading traffic generation systems use one or the other but their comparison has not yet been made; for example, Tmix [Tmix] uses non-parametric modeling while Swing [VV09] employs parametrized input for traffic generation.
2. how does experimental outcome differ when using same input traffic data but using per-connection round trip times (Tmix) vs. discrete approximation modeling of round trip times for emulating RTTs in traffic generation?
3. what are the differences in network experimentation when using same input traffic data on a network simulator vs. a laboratory testbed?
4. statistically analyze two sets of network data – one comprising wired traffic and the other comprising wireless traffic. Compare the results of the statistical analyses, and draw inferences on how this would affect application modeling for traffic generation in network experimentation.

These four research projects, among others, will be provided as guidance to the students. Upon their initial survey and readings, the students may pick one of the above ideas or something else they find interesting, within this area of network experimentation. Hence, while the first component of writing the survey paper is the same for all three students in this CREU project, the second component of the honors thesis will vary slightly, depending on the students' motivations and interests and the specific research questions they decide to explore, while remaining within the broader category of network experimentation.

## **Research Methods and Background Study**

All of the honors projects will employ research methods that have been used successfully by the networking research group in this department for over a decade. The PI runs a large networking research lab and will provide all of the equipment needed for this project. The students will run experiments using the Tmix traffic generation system [HC06] developed at UNC. They will employ measurement and modeling techniques and well-known statistical methods, while answering specific questions about these research methods that remain unanswered.

The common, collaborative component involving a survey of network experimentation will provide the students with rich background knowledge of the field that will prepare them to work on their honors research projects.

The overarching goal of this project is to expose the students to research involving network experimentation, and to encourage them to think about and employ scientific methods in the process of conducting such research.

## **Impact on the goals of CREU: Increase diversity in graduate schools and Provide Positive Research Experience**

Our department of Computer Science was first established as a graduate program and remained so for many years. It has only been over a decade that we have a thriving and successful undergraduate program for a BS in Computer Science. While our university itself has a larger ratio of women to men, the department of Computer Science, like most other such departments, remains male dominated. In our continued efforts to attract more women, minorities, or others who would not otherwise consider a degree in Computer Science, we constantly strive to collaborate with other departments on campus, like Information Science, Linguistics, and the Cancer research center. We have also recently established a BA in Computer Science. And yet, the number of women in Computer Science remains very low at all levels, from the undergraduate and graduate levels to faculty and upper-level administrative positions.

This project requires some background experience and coursework such that juniors would not be ready for the challenge. Hence, we have identified three undergraduate women students who will be seniors in the year 2011-2012. They are all highly committed to their education in Computer Science and have demonstrated excellence through their performance in the program. Each of them is engaged this summer in an internship on exciting projects: Rebecca at Google, Stephanie at an e-commerce startup, and Wai at the Air Force Research Laboratory. While each of them is an excellent student, they are often in classes with very few other women. This CREU project will provide these three bright minds with a collaborative, non-competitive atmosphere that allows independent as well as group work. They will provide support for each other, thus learning from each other as well.

Through this CREU project, we wish to ensure that these three students experience the joy of research and are encouraged to pursue graduate education as a result. Through the process of closely supervised research assignments, and frequent one-on-one discussions, including mandatory weekly group meetings, the undergraduate students will develop an appreciation for the role of cutting edge research in the field. This project aims to demystify 'Research' and give hands-on experience to these undergraduate students so that they get a realistic experience of the delights as well as the frustrations of conducting research. The process of designing experiments, setting up and running those experiments, collecting data, analyzing the results, presenting them in a research symposium, and writing a research report will all serve as excellent experience for these students.

### **Student Activities and Responsibilities**

The three students will each spend 10 hours per week on this project. That will initially include reading research papers assigned to them. This will help them prepare a survey paper of their own about the state of the art in network experimentation. Eventually, their work will become more heavily concentrated in the laboratory, first setting up, and then running network experiments that will answer the proposed research questions that will form the basis of their honors thesis.

They will attend a mandatory, joint, hour-long weekly meeting throughout the Fall and Spring semesters with the two PIs to facilitate teaching, planning and mutual feedback. As per departmental requirements for an honors thesis, they will each register for 3 credit hours for COMP 396 in both Fall and Spring semesters, under the supervision of Dr. Kevin Jeffay. They

will be required to provide weekly updates and a semester report as part of this Independent Study course.

As per requirements of the CREU program, each student will maintain a weekly journal documenting their work and findings, which they will share during the weekly meetings. Additionally, they will collaborate on maintaining a common website for their project, which will include their collaborative work, and later have separate details on each of their honors thesis projects. They will each provide a mid-year progress report at the end of the Fall 2011 semester, and an end of the year summary of their work. At that time, the website will also include links to their honors theses reports, which are typically 15-20 page research reports.

With help from the faculty advisors, the students will identify appropriate venues to publish their work, and send two papers – the first will be a collaborative survey paper on network experimentation, and the second will be an individual paper based on their honors thesis. As part of their honors thesis requirements, each student will also make a 20-minute oral presentation of her research during the undergraduate research symposium held at the end of the Spring semester.

### **PI Activities and Responsibilities**

The PI, Dr. Kevin Jeffay, runs a large laboratory funded by grants from the National Science Foundation, industry sponsors, and departmental support. As Director of Undergraduate Studies, Dr. Jeffay has mentored and supervised other groups of undergraduate students, including an all-women group, on research projects conducted in his lab. Along with a co-PI, Diane Pozefsky, he was funded for a CREU project titled “An Empirical Study of Network-Based Intrusion Detection Schemes” in 2004-2005. Of the three students who were part of that group, one pursued graduate studies at UC, Berkeley, the second pursued graduate studies at Columbia University, and the third is an entrepreneur with a thriving business.

As with his graduate students, Dr. Jeffay will provide his undergraduate mentees complete and open access to the networking laboratory resources. He will participate in weekly meetings, supervise the students, and provide all necessary support to the co-PI in this project.

Additionally, Dr. Aikat will work closely with the students, one-on-one in the lab, guiding them and helping them on their projects. As a woman researcher in this field, she is passionate about providing support and mentoring to women undergraduate students. In the past, while she was a graduate student, she mentored undergraduate minority students, and has twice (Spring 2009 and Fall 2009) won the Graduate Mentor Support Award from the Office of Undergraduate Research at UNC-CH for “excellent work with undergraduate mentoring”.

The PIs will be responsible for conducting weekly meetings with the undergraduate students to facilitate teaching, planning and feedback. This will provide needed feedback for the students. The undergraduate students will be supervised closely, while also given the freedom and exploration that should be an inherent part of gaining research experience. The CREU project will additionally facilitate more one-on-one discussions between each student and the PIs, thus enabling the PIs to closely mentor and guide these undergraduate students toward graduate studies in Computer Science, in choosing the programs, applying to them, and finding good mentors for graduate study.

## Project Timeline

### Fall 2011:

September – October: Identify, read and discuss research papers in the context of network experimentation and experimental methods used in these papers. These papers will be selected from the leading conferences and journals: ACM SIGCOMM (Special Interest Group on Data Communications) conference, IMC (Internet Measurement Conference), the ACM CoNEXT (Conference on emerging Networking EXperiments and Technologies), and the CCR (Computer Communication Review) journal.

November – December: Summarize their findings in a paper and send to the Computer Communications Review journal for publication. Identify specific research questions, either from the recommended list or any questions they wish to explore within this area of network experimentation and traffic data analyses.

### Spring 2012:

January: setup network testbed and run calibration experiments for the research project. This will involve designing, setting up, and running experiments, followed by data collection and analyses. If the project is based on analyzing traffic datasets, then this time will be spent on data collection, and writing and testing programs/scripts for data analysis.

February – March: run the complete set of experiments for the research project. Again, this will involve designing, setting up, and running experiments, followed by data collection and analyses. For traffic data analysis projects, this time will be spent sorting and organizing the data, running the programs/scripts, and statistically analyzing the data.

April: write up the 15-20 page honors thesis report, and present the research at the department's open undergraduate research symposium.

## Budget

Our budget is based on option # 1 of the CREU funding options. We are requesting academic year stipends of \$3000 per student for their work during the academic year, and additional \$1500 to offset the total cost of sending all three students to a conference.

- |  |          |
|--|----------|
| 1. Academic year stipend for three underrepresented students     | \$ 9,000 |
| 2. Travel allowance to send one or more students to a conference | \$ 1,500 |

**Total request** **\$10,500**

We would like to send each of the three students to a conference or workshop, preferably together, so they can interact with graduate students and researchers in the field. We realize that \$1500 is not sufficient for this effort; so we plan to seek funding from other venues, including the conference travel grant programs, and the Office of Undergraduate Research at UNC.

## **Role of this CREU project within the larger scope of this research / pedagogical project**

The first component of this CREU project will directly serve the pedagogical portion of the larger research project which involves teaching a course (and eventually writing a textbook) in experimental methods in networking. The survey of the state of the art in the field of network experimentation will provide a module to teach this class. This class is envisioned as an upper undergraduate / first-year graduate level research methods class in networking. Such a class is not being offered in leading universities, and the idea has received positive feedback from the educational community, including the leading SIGCOMM Education Workshop to be held in Toronto, Canada, in August 2011.

The second component is part of an ongoing research project that attempts to gain a better understanding of Internet traffic characteristics and the role of traffic generation techniques, the impact of application modeling and experimental methods on the outcome of network experimentation.

Both components of this CREU project, however, are designed to be independent self-contained projects, and targeted to providing a positive, rich, and realistic research experience for the undergraduate students conducting this project.

The PIs are also planning to offer a workshop with NSF support (funding is currently being sought for this) during Spring 2012. If this workshop is held, then the three students would participate directly in attending this workshop and interacting with other participants who will include leading researchers in the field as well as their graduate students.

## **Complete transcript for each student to be funded**

Please see pages 8 onwards.

## **References**

- [FP01] S. Floyd and V. Paxson. Difficulties in simulating the internet. *IEEE/ACM Transactions on Networking*, vol. 9, no. 4, pp. 392–403, August 2001.
- [HC06] Felix Hernandez-Campos. Generation and Validation of Empirically-Derived TCP Application Workloads. Ph.D. thesis, Univ. of North Carolina at Chapel Hill, Aug 2006.
- [NS] NS project. The Network Simulator—ns-2, <http://www.isi.edu/nsnam/ns/>
- [Tmix] <http://netlab.cs.unc.edu/Tmix>
- [VV09] K. Vishwanath and A. Vahdat, *Swing: Realistic and responsive network traffic generation*, *IEEE/ACM Transactions on Networking*, August 2009.