Khaled A. Harfoush – Curriculum Vitae

Department of Computer Science
Boston University
111 Cummington Street
Boston, MA 02215

Tel (Office): (617) 353-5222
(Home): (508)730-1229 & (508)353-4640
Email: harfoush@cs.bu.edu
WWW: http://cs-people.bu.edu/harfoush

INTERESTS

My main research interest is to develop end-to-end network diagnosis techniques to uncover dynamic network properties (e.g. congestion information, bottleneck equivalence, loss rates, topology, etc.) in a near real-time fashion. This information can be used to characterize the Internet dynamics and to dynamically adapt the control strategies of massively accessed Web servers (e.g. congestion control, redirection strategies, etc.) for a more efficient utilization of network resources.

EDUCATION

BOSTON UNIVERSITY — Boston, MA
Ph.D. candidate in Computer Science, degree expected in May 2002

CLEMSON UNIVERSITY — Clemson, SC
Ph.D. candidate in Computer Science

ALEXANDRIA UNIVERSITY — Alexandria, Egypt
M.Sc. Computer Science, June 1995

ALEXANDRIA UNIVERSITY — Alexandria, Egypt
B.Sc. Computer Science, June 1992 (distinction with honors)

DISSERTATIONS

Ph.D. Dissertation — **Advisor:** Prof. Azer Bestavros

A Framework For End-to-end Characterization of Metric-Induced Network Topologies: My Ph.D. research addresses the challenge of providing efficient, accurate and compact network models that are capable of uncovering shared network resources across multiple end-points. In my Ph.D. dissertation I introduced a framework that infers, labels and integrates *metric-induced network topologies* connecting network end-points. The framework applies to a broad class of metrics encompassing loss rate, delay, bandwidth … I also introduced a new set of end-to-end unicast probing mechanisms that are capable of uncovering different network properties. These mechanisms correctness have been validated through analysis, simulation and Internet deployment. They have been shown to be of great benefit for network-aware applications and for Internet characterization. (**Co-advisor:** Prof. John Byers)

Masters' Thesis — **Advisor:** Prof. Amin Shoukry

Speech Recognition Using Neural Networks: In my masters' thesis I developed a hierarchy of time-delay Neural Networks and single layer *Perceptrons* that were used for speech recognition.

RESEARCH EXPERIENCE

BOSTON UNIVERSITY — Boston, MA
Research Fellow — Fall 1998 to present
Advisor: Prof. Azer Bestavros
Designed, analyzed and implemented a framework (MINT) that is capable of diagnosing Internet dynamics. The MINT framework has been used as part of the MASS and WING networking groups at Boston University to analyze and characterize Internet performance and to design content delivery protocols for

massively accessed web servers. This work resulted in papers at INFOCOM'02, PAM'02, ICNP'00 and a couple of other papers under review.

CLEMSON UNIVERSITY Clemson, SC
Research Assistant Spring 1996 to Fall 1997
Advisor: Prof. Roy Pargas
I was responsible for implementing a Balanced Inventory Flow Replenishment System (BIFRS) for the Department of Defense. BIFRS is an enterprise-wide constraint management solution to optimize items distribution on time and at lowest total cost.

ALEXANDRIA UNIVERSITY Alexandria, Egypt
Research Assistant Fall 1992 to Summer 1994
Advisor: Prof. Amin Shoukry
Experimented with several *supervised* and *unsupervised* Neural Network models for the speaker independent recognition of speech utterances. Different Neural Network objective functions, methods of features extraction and ways of end-pointing speech utterances have been investigated. A new hierarchy consisting of modules of Time-Delay Neural Networks (TDNN) and Single Layer Perceptrons (SLP) has been proposed, analyzed and implemented. This work resulted in a paper at NSRC'95.

TEACHING EXPERIENCE

BOSTON UNIVERSITY Boston, MA
Computer Science Department
Teaching Assistant Fall 1997 to Fall 1998
Introduction to Computers (CS101)

BOSTON UNIVERSITY Boston, MA
Metropolitan School
Lecturer Spring 2000
Object Oriented Programming (METCS673)

BOSTON UNIVERSITY Boston, MA
School of Management
Teaching Assistant Present
Introduction to Networked Information Systems (IS 323 X3)

ARAB ACADEMY FOR SCIENCE AND TECHNOLOGY Alexandria, EGYPT
Department of Electronics and Computer Engineering
Lecturer Fall 1994 to Summer 1997
I have been an instructor for several undergraduate and graduate courses. These include Introduction to Computers, Structured Programming, Data Structures, Object Oriented Programming, Database Design, Database Applications, Numerical Analysis, Computer Graphics and Neural Networks. I have also supervised many undergraduate and graduate student projects. These projects dealt primarily with Neural Networks, Fuzzy Systems and Pattern Recognition.

AMERICAN UNIVERSITY IN CAIRO Cairo, EGYPT
Department of Continuing Education
Lecturer Fall 1993 to Summer 1997
I have been an instructor mainly for Database and Graphics courses.

PUBLICATIONS

- Bestavros, Azer; Byers, John; Harfoush, Khaled. *Inference and Labeling of Metric-Induced Network Topologies*. To appear in Proceedings of IEEE INFOCOM 2002, New York City, New York, June 2002.

- Harfoush, Khaled; Bestavros, Azer; Byers, John. *PeriScope: An Active Probing API*. To appear in Proceedings of PAM 2002, Passive and Active Measurement Workshop, Fort Collins, Colorado, March 2002.

- Harfoush, Khaled; Bestavros, Azer; Byers, John. *Measuring Bottleneck Bandwidth of Targeted Path Segments*. Technical Report BUCS-TR-2001-016 and submitted to ACM SIGCOMM 2002 for publication.

- Harfoush, Khaled; Bestavros, Azer; Byers, John. *Unicast-based Characterization of Network Loss Topologies*. In Proceeding of ICNP 2000: The 6th IEEE International Conference on Network Protocols (ICNP), Osaka, Japan, October 2000.

- Shoukry, Amin; El-Shehaby, Saleh; Khalaf, Aiman; Harfoush, Khaled. *Neural Network Classification of Remotely Sensed Images*. In Proceedings of NSRC'95 the twelfth National Radio Science Conference, Cairo, Egypt, March1995.

Most of these papers are electronically available off my web page: http://cs-people.bu.edu/harfoush. Also, most of these papers (as well as earlier versions thereof) have been published as BUCS technical reports, which can be found at http://www.cs.bu.edu/techreports.

PROFESSIONAL SERVICE

Program coordinator of WING'01 (the 2001 Workshop of the Web and Inter-Networking Group at Boston University), October 2001.

Referee for submissions to the following professional conferences and journals: INFOCOM'02 and IEEE Transactions on Networking.

REFERENCES

Prof. Azer Bestavros
Department of Computer Science
Boston University
Boston, MA 02215
E-mail: best@cs.bu.edu

Prof. John Byers
Department of Computer Science
Boston University
Boston, MA 02215
E-mail: byers@cs.bu.edu

Prof. Ibrahim Matta
Department of Computer Science
Boston University
Boston, MA 02215
E-mail: matta@cs.bu.edu

Prof. Mark Crovella
Department of Computer Science
Boston University
Boston, MA 02215
E-mail: crovella@cs.bu.edu

Khaled A. Harfoush – Research Statement

I am strongly driven by the desire to produce research with long lasting impact and strongly believe that good researchers should not only know how to approach a problem at hand but should also know how to spot important problems in the field.

My general areas of research interest are in Networking—a constantly evolving field with ample research opportunities. My previous research projects have focused on two complementary problems. The first concerned the end-to-end characterization of Network properties through the construction of compact, efficient Network models that capture Network dynamics. The availability of such models would assist in the development of network-aware services. The second problem concerned the adaptation of control strategies of transport protocols and network applications at massively accessed Internet servers in order to more efficiently utilize shared network resources and optimize the content delivery process.

My initial approach to solving these (as well as other) problems typically involved the use of modeling and mathematical analysis to generate exact results. This is followed by the use of more empirical approaches involving implementation and deployment to enable experimental evaluation under realistic conditions and assumptions.

1. Overview

Most applications nowadays do not provide service guarantees and thus hosts may experience varying performance over the lifetime of a connection. Network-aware applications attempt to react to changes in network resource availability and/or network performance. This reaction to network conditions is essential to better utilize network resources and optimize content delivery, especially because of the considerable strain on Internet resources imposed by the phenomenal growth of the World Wide Web. In order for Network-aware applications to function properly they need accurate, efficient and scalable models capturing network conditions and properties over *interesting* parts of the network.

Massively Accessed Scalable Servers (a.k.a. Mass servers) are popular Internet servers, which produce a substantial fraction of the traffic flowing through the network. Mass servers are uniquely positioned (1) to observe and diagnose network conditions by tracking the flows that they generate, and (2) to manage and control network resources by better regulating and scheduling the traffic they inject into the network

It is desirable to achieve these goals over a wide spectrum of time scales. Over shorter time scales, a Mass server can minimize packet loss by smoothing the (bursty) process of injecting packets into the network. Over longer time scales, a Mass server can perform aggregate congestion control by wisely bundling like connections to avoid the burstiness that results from competition among flows.

Network diagnostic models can play a big role in Internet characterization. They can also optimize the deployment of a variety of applications and services. Examples include Server Selection, Overlay Network Organization, Admission Control, Flow Scheduling and Cache/Replica Placement. The various implications and benefits of Network models provide a large pool of research opportunities and promise significant research impacts.

I have made a number of research contributions related to the above general goals. My recent contributions have been in the areas of Internet measurements and diagnosis, Network modeling and management, and control protocols and services. For example:

- I introduced the *MINT framework* to efficiently and compactly model network properties between end-points. The framework provides mechanisms for the *integration* of models obtained at different points in time or space.
- I proposed and implemented new *unicast end-to-end* probing techniques that enable the correlation of observations collected from end-points and a probing technique.
- I proposed and implemented a novel probing technique (called *Cartouche Probing*) that can efficiently infer the *bottleneck bandwidth of a path prefix.*

- I implemented a *generic active probing toolkit* (PeriScope). The toolkit is implemented in the *Linux* kernel, implements the MINT framework functionality and is capable of managing general user-defined probing structures.

Prior to my work on networking related problems, I had also worked on various problems in the field of Neural Networks and Speech Recognition. A full list of my publications is electronically available off of my Web page (http://cs-people.bu.edu/harfoush).

In the following sections I describe three representative pieces of work that I have done. In each section I will highlight the problem, describe the solution approach and the impact of my research.

2. Metric-Induced Network Topologies

Problem: The development and deployment of distributed network-aware applications and services over the Internet require the ability to compile and maintain a model of the underlying network resources with respect to (one or more) characteristic properties of interest. To be manageable, such models must be compact, and must enable a representation of properties along temporal, spatial, and measurement resolution dimensions.

Approach: We proposed a general framework for the construction of such metric-induced models using end-to-end measurements. We instantiated our approach using one such property, packet loss rates, and present an analytical framework for the characterization of Internet loss topologies. From the perspective of a server the loss topology is a logical tree rooted at the server with clients at its leaves, in which edges represent lossy paths between a pair of internal network nodes. We show how end-to-end unicast packet probing techniques could be used to (1) infer a loss topology and (2) identify the loss rates of links in an existing loss topology. We report on simulation, implementation, and Internet deployment results that show the effectiveness of our approach and its robustness in terms of its accuracy and convergence over a wide range of network conditions. A contribution of this work is to provide a mechanism to integrate metric-induced models collected at different hosts into one larger model. Another contribution is to provide a mechanism to integrate different metric-induced models collected from the same host at different points in time. These integration mechanisms allow the uncovering of more network details.

Output and Impact: This work has resulted in two publications. The framework itself is described in a paper to appear in INFOCOM'02 [1] and the framework implementation in the Linux kernel (a.k.a. the Periscope Toolkit) is described in a paper to appear in PAM'02 [2]. Periscope is being used by a number of researchers investigating various network-aware Internet and Peer-to-Peer applications.

3. End-to-end Characterization of Shared Loss Rates

Problem: Current Internet transport protocols make end-to-end measurements and maintain per-connection state to regulate the use of shared network resources. When two or more such connections share a common endpoint, there is an opportunity to correlate the end-to-end measurements made by these protocols to better diagnose and control the use of shared resources.

Approach: We developed packet-probing techniques to determine whether a pair of connections experience shared congestion. Our extensive simulation results demonstrated that the conditional (Bayesian) probing approach we employ provides superior accuracy, converges faster, and tolerates a wider range of network conditions than recently proposed memoryless (Markovian) probing approaches [5] for addressing this opportunity.

Output and Impact: This work appeared in the proceedings of ICNP'00 [4] and the Bayesian probing techniques are now part of the Periscope toolkit.

4. Bottleneck Bandwidth Along Targeted Path Segments

Problem: Accurate measurement of network bandwidth is crucial for flexible Internet applications and protocols, which actively manage and dynamically adapt to changing utilization of network resources. These applications must do so to perform tasks such as distributing and delivering high-bandwidth media, scheduling service requests and performing admission control. Extensive work has focused on two approaches to measuring bandwidth: measuring it hop-by-hop, and measuring it end-to-end along a path. Unfortunately, best-practice techniques for the former are inefficient and techniques for the latter are only able to observe bottlenecks visible at end-to-end scope.

Approach: We developed and simulated end-to-end probing methods, which can measure bottleneck bandwidth along arbitrary, targeted sub-paths of a path between two end-points in the network (including sub-paths shared by a set of flows). As another important contribution, we described a number of practical applications which we foresee as standing to benefit from solutions to this problem, especially in emerging, flexible network architectures such as overlay networks, ad-hoc networks, peer-to-peer architectures and massively accessed content servers.

Output and Impact: This work is submitted for publication to SIGCOMM'02 [3] (and is available as a Technical Report).

5. Future Research Directions

As a follow-up of my thesis research, there are many directions that I would like to explore further. Let me briefly discuss some of these directions:

- *Active vs. passive measurements:*
  Investigate whether it is possible to diagnose network conditions by simply inspecting original flow packets, without injecting additional probe packets in the network. For which metrics is this possible? How accurate are the results?

- *Other metric properties:*
  The MINT framework abstracts any metric based on three properties: Monotonicity, Separability and Symmetry. Are there other properties that can be exploited? If yes then what is their impact on the framework inference, labeling and integration procedures?

- *How best to exert control over flows given diagnosis?*
  Different applications need different views of the network (different metrics of interest and different diagnostic resolutions, etc). Given an appropriate network diagnosis, what is the optimal way for an application to use the information?

- *Bottleneck Bandwidth along arbitrary path segments:*
  Cartouche Probes are used to infer the Bottleneck Bandwidth over a path prefix. Investigate ways to extend Cartouche probes to estimate the Bottleneck Bandwidth along arbitrary path segments.

In addition to the above thesis-inspired problems, I am also very interested in pursuing research in other areas of networking research

- *Mobile ad-hoc networks:*
  Wireless information access is becoming increasingly important which leads to a growing interest in wireless mobile ad-hoc networks. The highly dynamic nature of these networks, the potentially complex communication environment and the varying capability needs of the participants limit traditional Internet protocols ability to manage mobile ad-hoc network resources. This research field s very promising offering a wide range of research opportunities ranging from assigning participants unique IP addresses to security implications. It has been planned that I will be installing wireless lab in

the school of Management at Boston University in the Spring 2002 semester. The lab will serve as a test bed for various wireless experiments.

- *End-to-end admission control:*
  In order to provide Quality of Service, the *Integrated* Services paradigm provides per flow guarantees but suffers scalability problems due to the need to store per flow state at core routers. On the other hand, the *Differentiated* Services paradigm tries to solve scalability issues by removing per flow state from core routers. Traffic entering the network is classified and conditioned at the network boundaries and packets are assigned service bits. Core routers inspect packets service bits and schedule packets accordingly providing guarantees to flow aggregates and not to flows. Both *Integrated* and *Differentiated* services need router support. End-to-end admission control looks into providing statistical guarantees to flows without router support by relying on network diagnosis information. The field is relatively new and needs a lot of investigation. It has the potential of being deployed in today's Internet.

While I am eager to establish my identity as an independent researcher, I believe that research is a collaborative effort. It is collaborative on many fronts between industry, funding agencies and academia (faculty members and students). As a future academic, I will strive to bring research parties together, bring to the table many research ideas and influence my students to be sharp and independent thinkers.

References

[1] Bestavros, Azer; Byers, John; Harfoush, Khaled. *Inference and Labeling of Metric-Induced Network Topologies*. To appear in Proceedings of IEEE INFOCOM 2002, New York City, New York, June 2002.

[2] Harfoush, Khaled; Bestavros, Azer; Byers, John. *PeriScope: An Active Probing API.* To appear in Proceedings of PAM 2002, Passive and Active Measurement Workshop, Fort Collins, Colorado, March 2002.

[3] Harfoush, Khaled; Bestavros, Azer; Byers, John. *Measuring Bottleneck Bandwidth of Targeted Path Segments*. Technical Report BUCS-TR-2001-016 and submitted to ACM SIGCOMM 2002 for publication.

[4] Harfoush, Khaled; Bestavros, Azer; Byers, John. *Unicast-based Characterization of Network Loss Topologies*. In Proceeding of ICNP 2000: The 6th IEEE International Conference on Network Protocols (ICNP), Osaka, Japan, October 2000.

[5] D. Rubenstein, J. Kurose and D. Towsley. Detecting Shared Congestion of Flows via End-to-end Measurement. In Proceedings of ACM SIGMETRICS'00, Santa Clara, CA, June 2000.

Khaled A. Harfoush – Teaching Statement

My strong belief in the important role of teaching fuels my motivation to pursue a career in academia. I believe that teaching complements research in the academic wheel and thus a teaching strategy should establish a close relationship between research and education. The teacher, as a guide to his or her students, has the responsibility not only to providing support and help to his students in the areas in which help is most needed, but also to making teaching an interactive, enjoyable experience, encouraging students to think independently and critically. Students should not only understand the material, but they should also understand how the material is to benefit them in the future.

As detailed in my Curriculum Vitae, I have a long experience in teaching at Boston University, the Arab Academy for Science and Technology and the American University in Cairo, both as a teaching assistant and as a lecturer.

I have been a teaching assistant in many courses, including Introduction to Computers, Introduction to Information Systems, Numerical Analysis and Computer Graphics. My responsibilities in these courses included, in addition to teaching, participating in the development of homework assignments, projects, midterms and finals. It also included grading and mentoring students in various projects.

I have been the main instructor in a number of other courses, including Introduction to Computers, Structured Programming, Data Structures, Object Oriented Programming, Database Design, Database Applications, Computer Graphics and Neural Networks. I found the experience of independently teaching and managing all the details of these courses to be very rewarding. I realized that careful planning (e.g., making the learning resources such as lecture notes and references to supplementary reading easily accessible) makes the teaching experience enjoyable both for the students and the teacher. I also realized that my research greatly informs my teaching. Doing research in a subject related to a course I am teaching allows me to provide students with deeper insights into the course material—not surprisingly, students recognize and appreciate this duality.

I have been a lecturer in the department of Electronics and Computer Engineering at the Arab Academy for Science and Technology at the time they were re-examining, re-designing and standardizing all their courses in order to be certified by ISO as a degree-granting Computer Engineering department. As a member of the technical committee in charge of this job, I helped developing the curricula and the details of some computer science courses. These courses were Object Oriented Programming, Structured Programming, Data Structures and Neural Networks.

My computer science and engineering background allows me to teach a wide variety of courses ranging from theory to systems to programming languages. Given my research interests, I would be particularly interested (and indeed excited) to teach introductory and advanced courses in Networking, Distributed Systems and Performance Evaluation. I would also enjoy giving seminars on graduate level topics like End-to-end Characterization, Network-aware Applications, Quality of Service and Mobile Ad-hoc Networks.

I have specific ideas concerning Computer Science courses and seminars that I would like to develop further as part of my future teaching career. For example:

- In order to boost Internet measurements and characterization research I propose to either develop new courses on Internet Measurements or include aspects of Internet Measurement in already existing courses. This would allow describing the various techniques and tools involved in taking Internet measurements, describing the subtleties involved and ways of filtering spurious measurements.

- In order to satisfy the needs of graduate students at different stages of their study; introduce Networking seminar courses and reading groups at two different levels of detail. The first is about networking in general, while the second is dedicated for more specific networking subjects (e.g. QoS, wireless communications, Internet measurements, etc…). The general seminars introduce students to the different areas of networking research, while the more specific seminars provide them with more details on areas they are mostly interested in. These seminars can lead to new courses over time.

Computer Science is continuously evolving and broadening. Given the limited number of courses an undergraduate student is expected to take, it is important to keep CS curricula focused on fundamental concepts that survive technological trends. To do so requires a continuous critical look at emerging areas in CS in order to identify and distill those fundamental concepts that are worthy of inclusion in an undergraduate curriculum. Thus, as a CS educator, I will strive not only to excel in teaching a specific set of courses, but also to follow the most recent trends and ideas in CS in general, and in networking in particular, with an eye on adapting the curriculum accordingly. This is a challenge I look forward to undertake.