

Raja R. Sambasivan

31 Sunset Ave
Medford, MA 02155
412-983-1701
raja AT cs.tufts.edu
<http://www.rajasambasivan.com>

Tufts University
Computer Science Department
161 College Avenue
Medford, MA 02155

EXPERIENCE

Ankur & Mari Sahu Assistant Professor *August 2019 – Present*
Tufts University, Computer Science Department

Research Scientist *November 2016 – July 2019*
Boston University, Mass Open Cloud & Red Hat Collaboratory

Postdoctoral Researcher *June 2013 – September 2016*
Carnegie Mellon University, eXpressive Internet Architecture (XIA) Group

Consultant *November 2014 – December 2014*
Huawei, storage group

PhD Student *June 2006 – May 2013*
Carnegie Mellon University, Parallel Data Lab

Software Engineering Intern *May 2010 – December 2010*
Google

Research Intern *July 2007 – March 2008*
HP Labs

Systems Programmer *June 2004 – May 2006*
Carnegie Mellon University, Parallel Data Lab

EDUCATION

Ph.D., Electrical & Computer Engineering, May 2013
Carnegie Mellon University, Pittsburgh, PA
Advisor: Greg Ganger
Dissertation: Diagnosing performance changes in distributed systems by comparing request flows

M.S., Electrical & Computer Engineering, May 2004
Carnegie Mellon University, Pittsburgh, PA

B.S., Electrical & Computer Engineering w/minor in Computer Science, May 2003
Carnegie Mellon University, Pittsburgh, PA

HONOURS & AWARDS

Ankur & Mari Sahu Endowed Assistant Professorship, August 2019.

Best poster, EMC University Day 2012 (*Diagnosing performance changes by comparing request flows*).

Best paper, SIGMETRICS 2007 (*Modeling the relative fitness of storage*).

Featured in Piled Higher & Deeper. ([PhDComics](#)), February 14th, 2007. [strip](#).

Best paper, FAST 2005 (*Ursa Minor: versatile cluster-based storage*).

TEACHING
EXPERIENCE

Course developer & instructor
COMP/CS 118, Cloud Computing

Fall 2021, Fall 2020, Spring 2020

This class covers foundational cloud-computing concepts and technologies. The curriculum includes lectures on research papers, homeworks, and projects within a major cloud platform.

Enrollment: F'20: 27; S'20: 37,

Overall rating (out of 5): F'20: 3.96; S'20: 3.52

Instructor rating (out of 5): F'20: 4.12; S'20: 3.98

Course developer & instructor

Spring 2021, Fall 2019

COMP 150-DCC, *Special Topics: Debugging Cloud Computing*

This class introduces students to state-of-the-art techniques for diagnosing correctness and performance problems in distributed systems and networks. The curriculum emphasizes reading and discussing research papers on logging, distributed tracing, provenance, and network telemetry. Students are required to perform an in-depth, semester-long research project.

Enrollment: S'21: 17; F'19: 5

Overall rating (out of 5): S'21: 4.52; F'19: 4.35

Instructor rating (out of 5): S'21: 4.46; F'19: 4.71

Guest lecturer & project mentor

Spring 2017, Spring 2018 (guest lecturer only)

CS528, *Cloud Computing*

This class provides an overview of cloud-computing concepts via a curriculum that emphasizes reading research papers and semester-long projects. The class usually consists of 30-50 undergraduates and Master's students. I guest lectured on topics related to networking and problem diagnosis. In the Spring of 2017, I additionally mentored students on projects related to distributed tracing.

Course Co-developer & co-instructor

Fall 2013

CS 15-719, *Advanced Cloud Computing*

This class provides an overview of cloud-computing concepts via a curriculum that emphasizes reading research papers, lectures, projects, and exams. It consisted of about 30 Master's students and a few PhD students when I taught it.

Teaching Assistant

Fall 2005 & Spring 2010

ECE 18-746, *Storage Systems*

This class covers a broad range of material, including hard-disk architecture, file-system design and debugging, RAID, and object-based storage. It usually consists of 40-60 Master's students and PhD students.

UNIVERSITY
SERVICE

Tufts CS Department:

- Graduate committee F'20 - S'21
- Graduate student open-house co-organizer: F'20 - S'21, F'19 - S'20
- Graduate admissions committee: F'20 - S'21, F'19 - S'20
- Colloquium co-organizer: F'20 - S'21
- Software systems master's program development committee member: F'19 - S'20

PROFESSIONAL
SERVICE

Program committee member:

- ACM Symposium on Cloud Computing (SoCC) 2021
- IEEE International Conference on Autonomic Computing and Self-Organizing Systems (ACSOS) 2020
- IEEE Transactions on Services Computing 2019, 2015
- IEEE Transactions on Software Engineering 2017
- ACM HotStorage 2014, 2013

External reviewer: SIGMETRICS 2021, SIGCOMM CCR 2018, EuroSys 2017

Panel member: NSF 2020, NSF Cloud Review 2017, NSF 2016

Session chair: NENS 2019, NENS 2017, SoCC 2016

Reading groups: Tufts diagnosis (2020-present) BU diagnosis (2017-2019), CMU network diagnosis (2013-2014), CMU diagnosis (2011-2013)

Working groups: Red Hat telemetry co-lead (Fall 2020 - present)

PROSPECTUS,
PROPOSAL, AND
DISSERTATION
COMMITTEES

Mert Toslali, Boston University, July 2021, Prospectus Defense

Mania Abdi, Northeastern University, December 2020, Proposal defense

Title: Characterizing, debugging, and performance optimization of cloud applications using graph processing

Emre Ates, Boston University, Dissertation defense, June 2020

Title: Automating Telemetry- and Trace-Based Analytics on Large-Scale Distributed System

OUTREACH

Letter Writer, Letters to a Pre-Scientist middle-school outreach program (2018)

Mentor, MIT Primes high-school research program (2017-present)

- Mentees named Siemens Competition Semi-Finalists (2017)
- Mentees selected to present at Red Hat Developers' Conference Devconf.us (2018)
- Mentee awarded 2nd prize in MA state Science Fair (2020)

CS Grand Awards Judge, Intel Science & Engineering Fair Finals (2015, 2012)

Presenter, Carnegie Science Center Buhl Planetarium (2010)

MENTORING

Current advisees (in order of matriculation):

- Max Liu, PhD student, Tufts University (Fall 2020 -)
- James Mattei, PhD student, Tufts University (Fall 2020 -)
- Zhaoqi Zhang, PhD student, Tufts University (Fall 2020 -)
- Darby Huye, MS & PhD student, Tufts University (MS: Fall 2020, PhD: Spring 2021 -)
- Alexander Ellis, MS thesis student, Tufts University (Spring 2020 -)
- Tomislav Zabcic-Matic, PhD student, Tufts University (Summer 2021 -)
- Sarah Abowitz, PhD student, Tufts University (Fall 2021 -)

Current High-school students (in order of matriculation):

- Tanmay Gupta (Fall 2020 -)
- Anshul Rastogi (Fall 2020 -)

Other current students (in order of matriculation):

- Mania Abdi, PhD, Northeastern University (Fall 2018 -)
Advisor: Peter Desnoyers
- Mert Toslali, PhD, Boston University (Fall 2018 -)
Advisor: Ayse Coskun

Graduated students (most recent to least recent):

- Emre Ates, PhD, Boston University (Fall 2018 - Spring 2020)
Advisor: Ayse Coskun
Dissertation: Automating telemetry and trace-based analytics on large-scale distributed systems
Next step: Google
- Neel Bhalla, Lexington High School'20 (Spring 2018 - Spring 2020)

Next step: Northeastern University

- Lily Sturmann, MS, Harvard Extension School (Spring 2017 - Spring 2019)
Thesis: Using Performance Variation for Instrumentation Placement in Distributed Systems
Next step: Red Hat
- Harshal Sheth, High School, Westford Academy (Spring 2017 - Spring 2018)
Next step: Yale University
- Andrew Sun, High School, Westford Academy (Spring 2017 - Spring 2018)
Next step: Northeastern University

Other students I've worked with (most recent to least recent)

- Runze Si, MS, Boston University
- Golsana Ghaemi, PhD, Boston University
Advisor: Orran Krieger
- Emine Ugur Kaynar, PhD, Boston University
Advisor: Orran Krieger
- Jethro Sun, MS'18, Boston University
Advisor: Orran Krieger
- Da Yu, PhD, Brown University
Advisor: Rodrigo Fonseca
- David Tran-Lam, PhD, University of Wisconsin-Madison
Advisor: Aditya Akella
- William Wang, MS, Carnegie Mellon University
Advisor: Greg Ganger

Raja R. Sambasivan

FUNDING

A just-in-time, cross-layer instrumentation framework for diagnosing performance problems in distributed applications. Raja R. Sambasivan, Ayse K. Coskun, Orran Krieger. \$460,249. Award # [CNS-1815323](#). October 2018 to October 2021.

REFEREED PUBLICATIONS

Automating instrumentation choices for performance problems in distributed applications with VAIF. Mert Toslali, Emre Atex, Alexander Ellis, Zhaoqi Zhang, Darby Huye, Liu Lan, Samantha Puterman Ghitelman, Ayse K. Coskun, Raja R. Sambasivan. In Proceedings of the 12th ACM Symposium on Cloud Computing (SoCC'21). November 1st to November 3rd, 2021. To appear.

D3N: A multi-layer cache for the rest of us. Emine Ugur Kaynar, Mania Abdi, Mohammad Hossein Hajkazemi, Ata Turk, Raja R. Sambasivan, Larry Rudolph, Peter Desnoyers, Orran Krieger. In proceedings of the 2019 IEEE International Conference on Big Data (BigData'19). December 9th to December 12th, 2019. Los Angeles, CA.

An automated, cross-layer instrumentation framework for diagnosing performance problems in distributed applications. Emre Ates, Mert Toslali, Richard Megginson, Orran Krieger, Ayse K. Coskun, Raja R. Sambasivan. In Proceedings of the 10th ACM Symposium on Cloud Computing (SoCC'19). November 20th to November 23rd, 2019. Santa Cruz, CA.

Bootstrapping evolvability for inter-domain routing with D-BGP. Raja R. Sambasivan, David Tran-Lam, Aditya Akella, Peter Steenkiste. In proceedings of the ACM 2017 SIGCOMM Conference (SIGCOMM'17). August 21th to August 25th, 2017. Los Angeles, CA, USA.

Principled workflow-centric tracing of distributed systems. Raja R. Sambasivan, Ilari Shafer, Jonathan Mace, Rodrigo Fonseca, Gregory R. Ganger. In proceedings of the 7th ACM Symposium on Cloud Computing (SoCC'16). October 5th to October 7th, 2016. Santa Clara, CA, USA.

Bootstrapping evolvability for inter-domain routing. Raja R. Sambasivan, David Tran-Lam, Aditya Akella, Peter Steenkiste. In proceedings of the 14th ACM Workshop on Hot Topics in Networks (HotNets'15). November 16th to November 17th, 2015. Philadelphia, PA, USA.

Visualizing request-flow comparison to aid performance diagnosis in distributed systems. Raja R. Sambasivan, Ilari Shafer, Michelle L. Mazurek, Gregory R. Ganger. IEEE Transactions on Visualization and Computer Graphics 19(12), December 2013. In proceedings of Information Visualization 2013.

Specialized storage for big numeric time series. Ilari Shafer, Raja R. Sambasivan, Anthony Rowe, Gregory R. Ganger. In proceedings of the 5th USENIX Workshop on Hot Topics in Storage and File Systems (HotStorage'13). June 27th to June 28th, 2013. San Jose, CA, USA.

Automated diagnosis without predictability is a recipe for failure. Raja R. Sambasivan, Gregory R. Ganger. In proceedings of the 4th USENIX Workshop on Hot Topics in Cloud Computing (HotCloud'12). June 12th to June 13th, 2012. Boston, MA, USA.

Diagnosing performance changes by comparing request flows. Raja R. Sambasivan, Alice X. Zheng, Michael De Rosa, Elie Krevat, Spencer Whitman, Michael Stroucken, William Wang, Lianghong Xu, Gregory R. Ganger. In proceedings of the 8th USENIX Symposium on Network Systems Design and Implementation (NSDI'11). March 30th to April 1st, 2011. Boston, MA, USA.

A transparently-scalable metadata service for the Ursa Minor storage system. Shafeeq Sinnamohideen, Raja R. Sambasivan, Likun Liu, James Hendricks, Gregory R. Ganger. In proceedings of the 2010 USENIX Annual Technical Conference (USENIX ATC'10). June 23rd to 25th, 2010. Boston, MA, USA.

Categorizing and differencing system behaviours. Raja R. Sambasivan, Alice X. Zheng, Eno Thereska, Gregory R. Ganger. Appears in the proceedings of the 2nd International Workshop on Hot Topics in Autonomic Computing (HotAC II). June 15th, 2007. Jacksonville, Florida, USA.

Modeling the relative fitness of storage. Michael Mesnier, Matthew Wachs, Raja R. Sambasivan, Alice X. Zheng, Gregory R. Ganger. In proceedings of the International Conference on Measurement and Modeling of Computer Systems (SIGMETRICS'07). June 12th to 16th, 2007. San Diego, CA, USA.

//TRACE: parallel trace replay with approximate causal events. Michael Mesnier, Matthew Wachs, Raja R. Sambasivan, Julio Lopez, James Hendricks, Gregory R. Ganger. In proceedings of the 5th conference on File and Storage Technologies (FAST'07). February 13th to 16th, 2007. San Jose, CA, USA.

Ursa Minor: versatile cluster-based storage. Michael Abd-El-Malek, William V. Courtright II, Chuck Cranor, Gregory R. Ganger, James Hendricks, Andrew J. Klosterman, Michael Mesnier, Manish Prasad, Brandon Salmon, Raja R. Sambasivan, Shafeeq Sinnamohideen, John D. Strunk, Eno Thereska, Matthew Wachs, Jay J. Wylie. In the proceedings of the 4th USENIX conference on File and Storage Technologies (FAST'05). December 13th to 16th, 2005. San Francisco, CA, USA.

Replication policies for layered clustering of NFS servers. Raja R. Sambasivan, Andrew J. Klosterman, Gregory R. Ganger. Appears in the proceedings of the 13th Annual Meeting of the IEEE International Symposium on Modeling, Analysis, and Simulation of Computer and Telecommunication Systems (MASCOTS'05). September 27th to 29th, 2005. Atlanta, Georgia, USA.

JOURNAL
PUBLICATIONS

Relative fitness modeling. Michael Mesnier, Matthew Wachs, Raja R. Sambasivan, Alice Zheng, Raja R. Sambasivan, Gregory R. Ganger Research Highlights, Communications of the ACM. April 2009.

Early experiences on the journey towards self-* storage. Michael Abd-El-Malek, William V. Courtright II, Chuck Cranor, Gregory R. Ganger, James Hendricks, Andrew J. Klosterman, Michael Mesnier, Manish Prasad, Brandon Salmon, Raja R. Sambasivan, Shafeeq Sinnamohideen, John D. Strunk, Eno Thereska, Matthew Wachs, Jay J. Wylie. In the Bulletin of the IEEE Computer Society Technical Committee on Data Engineering 29(3). Special issue on self-managing database systems. September 2006.

TECHNICAL
REPORTS

Bootstrapping evolvability for inter-domain routing with D-BGP. Raja R. Sambasivan, David Tran-Lam, Aditya Akella, Peter Steenkiste. Carnegie Mellon Computer Science Technical Report CMU-CS-16-117. June 2016.

So, you want to trace your distributed system? Key design insights from years of practical experience. Raja R. Sambasivan, Rodrigo Fonseca, Ilari Shafer, Gregory R. Ganger. Carnegie Mellon University Parallel Data Laboratory Technical Report CMU-PDL-14-102. April 2014.

Visualizing request-flow comparison to aid performance diagnosis in distributed systems. Raja R. Sambasivan, Ilari Shafer, Michelle L. Mazurek. Carnegie Mellon University Parallel Data Laboratory Technical Report CMU-PDL-13-104. May 2013. Supersedes CMU-PDL-12-102.

Visualizing request-flow comparison to aid performance diagnosis in distributed systems. Raja R. Sambasivan, Ilari Shafer, Michelle L. Mazurek. Carnegie Mellon University Parallel Data Laboratory Technical Report CMU-PDL-12-102. May 2012.

Automation without predictability is a recipe for failure. Raja R. Sambasivan, Gregory R. Ganger. Carnegie Mellon University Parallel Data Laboratory Technical Report CMU-PDL-11-101. January 2011.

Diagnosing performance changes by comparing system behaviours. Raja R. Sambasivan, Alice X. Zheng, Elie Krevat, Spencer Whitman, Michael Stroucken, William Wang, Lianghong Xu, Gregory R. Ganger. Carnegie

Mellon University Parallel Data Laboratory Technical Report CMU-PDL-10-107. July 2010. Supersedes CMU-PDL-10-103.

A transparently-scalable metadata service for the Ursa Minor storage system. Shafeeq Sinnamohideen, Raja R. Sambasivan, James Hendricks, Likun Liu, Gregory R. Ganger. Carnegie Mellon University Parallel Data Laboratory Technical Report CMU-PDL-10-102. March 2010.

Diagnosing performance problems by visualizing and comparing system behaviours. Raja R. Sambasivan, Alice X. Zheng, Elie Krevat, Spencer Whitman, Gregory R. Ganger. Carnegie Mellon University Parallel Data Lab Technical Report CMU-PDL-10-103. February 2010.

Eliminating cross-server operations in scalable file systems. James Hendricks, Shafeeq Sinnamohideen, Raja R. Sambasivan, Gregory R. Ganger. Carnegie Mellon University Parallel Data Lab Technical Report CMU-PDL-06-105. May 2006.

Improving small file performance in object-based storage. James Hendricks, Raja R. Sambasivan, Shafeeq Sinnamohideen, Gregory R. Ganger. Carnegie Mellon University Parallel Data Lab Technical Report CMU-PDL-06-104. May 2006.

Selected project reports, Spring 2005 Advanced OS & Distributed Systems (15-712). Garth A. Gibson and Hyang-Ah Kim, Editors. Jangwoo Kim, Eriko Nurvitadhi, Eric Chung; Alex Nizhner, Andrew Biggadike, Jad Chamcham; Srinath Sridhar, Jeffrey Stylos, Noam Zeilberger; Gregg Economou, Raja R. Sambasivan, Terrence Wong; Elaine Shi, Yong Lu, Matt Reid; Amber Palekar, Rahul Iyer. Carnegie Mellon Computer Science Technical Report CMU-CS-05-138. May 2005.

Ursa Minor: Versatile cluster-based storage. Michael Abd-El-Malek, William V. Courtright II, Chuck Cranor, Gregory R. Ganger, James Hendricks, Andrew J. Klosterman, Michael Mesnier, Manish Prasad, Brandon Salmon, Raja R. Sambasivan, Shafeeq Sinnamohideen, John D. Strunk, Eno Thereska, Matthew Wachs, Jay J. Wylie. Carnegie Mellon University Parallel Data Laboratory Technical Report CMU-PDL-05-104. April 2005.

PATENTS

Managing execution of database queries. Stefan Kompers, Harumi Anne Kuno, Umeshwar Dayal, Janet Wiener, Raja Sambasivan. U.S. Patent 9,910,892. March 6th, 2018.

CONFERENCE TALKS

Bootstrapping evolvability for inter-domain routing with D-BGP. Presented at the ACM 2017 SIGCOMM Conference (SIGCOMM'17). August, 2017.

Principled workflow-centric tracing of distributed systems. Presented at the 7th ACM Symposium on Cloud Computing (SoCC'16). October 2016.

Bootstrapping evolvability for inter-domain routing. Presented at the 14th ACM Workshop on Hot Topics in Networks (HotNets'15). November 2015.

Visualizing request-flow comparison to aid performance diagnosis in distributed systems. Presented at IEEE InfoVis 2013.

Automated diagnosis without predictability is a recipe for failure. Presented at the 4th USENIX Workshop on Hot Topics in Cloud Computing (HotCloud'12). June 2012.

Generalizing request-flow comparison to more systems. WiP talk at 23rd ACM Symposium on Operating Systems Principles (SOSP'11). October 2011.

Diagnosing performance changes by comparing request flows. Presented at the 8th USENIX Symposium on Networked Systems Design and Implementation (NSDI'11). March 2011.

Spectroscope: a tool for categorizing and differencing system behaviours. Presented at the 2nd International Workshop on Hot Topics in Autonomic Computing (HotACII). June 2007.

Replication policies for layered clustering of NFS servers. Presented at the 13th Annual Meeting of the IEEE International Symposium on Modeling, Analysis, and Simulation of Computer and Telecommunication Systems (MASCOTS'05). September 2005.

INVITED TALKS &
GUEST LECTURES

Workflow-centric tracing and advanced diagnosis tools for the cloud ecosystem. Presented at UMass Amherst (February 2019), George Washington University (March 2019), Tufts University (March 2019).

Toward a diagnosis plane for cloud computing. Presented at LightStep (April 2018), Columbia University (March 2018), Facebook (February 2018), Brown University (February 2018).

Diagnosis and inter-domain support for an Internet of clouds. Presented at Tufts University (October 2017), Yale University (March 2017), AT&T Labs (May 2016), Intel Labs (April 2016), NYU (April 2016).

Diagnosing performance changes by comparing request flows. Presented at UCSD (April 2014), Brown University (April 2012), NetApp RTP (September 2011), Google NYC (June 2011).

Networking at Google: B4 & Jupiter Rising. Guest lecture in BU CS 528 (March 2018, March 2017).

When the cloud fizzles: Outages and how to debug them. Guest lecture in NU CS 6620 (April 2018) and BU CS 528 (April 2017).

A case study of the AWS outage on April 21st, 2011. Guest lecture in CMU 15-719 (Fall 2015, Fall 2014).

Diagnosis via monitoring & tracing. Guest lecture in CMU 15-719 (Fall 2015, Fall 2014).