BINGHAMTON
U N I V E R S I T Y | Computer Science
STATE UNIVERSITY OF NEW YORK

**THE DEPARTMENT OF COMPUTER SCIENCE & THE COMPUTER SCIENCE
GRADUATE STUDENT ORGANIZATION (GSOCS) PRESENT**

INVITED SPEAKER SERIES

Professor Emery Berger
University of Massachusetts at Amherst

Friday, December 11th at 12 noon
Zoom URL: https://binghamton.zoom.us/j/94028873916

Performance (Really) Matters

Abstract: Performance clearly matters to users. For example, the most common software update on the AppStore is "Bug fixes and performance enhancements." Now that Moore's Law has ended, programmers have to work hard to get high performance for their applications. But why is performance hard to deliver?

I will first explain why current approaches to evaluating and optimizing performance don't work. I'll show how complicated performance has become on modern systems, and how compiler optimizations have essentially run out of steam. Next, I'll introduce two radically new performance profilers that guide programmers directly to the code they need to change to improve application performance. The first is Coz, a new "causal profiler" for C/C++/Rust (with a Java version as well) that lets programmers optimize for throughput or latency, and which pinpoints and accurately predicts the impact of optimizations via what we call "virtual speedup" experiments. Coz's approach unlocks previously unknown optimization opportunities. Guided by Coz, we improved the performance of applications by as much as 68%; in most cases, this involved modifying less than 10 lines of code and took under half an hour (without any prior understanding of the programs!). Coz now ships as part of standard Linux distros (apt install coz-profiler).

The second is Scalene, a "scripting-language aware" profiler for Python (with other scripting languages to come). Scalene runs orders of magnitude faster than other profilers while delivering far more detailed information -- information that's especially valuable to Python programmers. Via a combination of sampling, inference, and disassembly of byte-codes, Scalene efficiently and precisely attributes execution time and memory usage to Python, which developers can optimize, or library code, which they cannot. Its novel sampling memory allocator efficiently reports line-level memory consumption and trends with low overhead, helping developers reduce footprints and identify leaks. Finally, Scalene reports a new metric, copy volume, that helps developers root out insidious copying costs across the Python/library boundary, which can drastically degrade performance. Scalene is available on PyPi (pip install scalene).

Bio: Emery Berger is a professor at the University of Massachusetts Amherst, where he co-leads the PLASMA research group (https://plasma-umass.org). His research interests span programming languages, systems, security, and human-computer interaction. His group is well known for its influential research and software systems, many of which have enjoyed broad adoption: among others, their work has influenced the development of the Rust and Swift programming languages, and memory managers deployed in both Mac OS X and Microsoft Windows. He also developed and maintains the CSrankings.org website. He was named an ACM Fellow in 2019.

https://www.emeryberger.com