

The University of Chicago Computer Science Department  
**PRESENTS:**

**Rethinking the Hardware-Software contract: Enabling practical and  
general cross-layer optimizations**



**Nandita Vijaykumar**  
Carnegie Mellon University

**Abstract:**

Layered abstractions in the computing stack are critical to building complex systems, but the existing \*interfaces\* between layers restrict what can be done at each level. Enhancing cross-layer interfaces--specifically, the hardware-software interface--is crucial towards addressing two important and hard-to-solve challenges in computer systems today: First, significant effort and expertise are required to write high-performance code that harnesses the full potential of today's diverse and sophisticated hardware. Second, as a hardware or system designer, architecting faster and more efficient systems is challenging as the vast majority of the program's semantic content and programmer intent gets lost in translation with today's hardware-software interface. Moving towards the future, these challenges in programmability and efficiency will be even more intractable as we architect increasingly heterogeneous and sophisticated systems.

In this talk, I will highlight my work [ISCA'15, MICRO'16, ISCA'18, ISCA'18] on how to design rich cross-layer abstractions that provide layered interfaces to directly communicate higher-level program semantics and intent from the application to the lower levels of the stack. In doing so, we can effectively bridge the so-called "semantic gap" between applications and computer systems, and enable a wide range of cross-layer optimizations in future systems with a single unifying interface. I will discuss how cross-layer approaches with these abstractions can significantly enhance (1) *performance and efficiency* by enabling the system to adapt to application characteristics and (2) *programmability and portability* by enabling application software to easily leverage diverse underlying hardware resources without specific knowledge of system details. For example, daunting aspects of programming GPUs can be made much simpler with a rich cross-layer programming abstraction. I will describe how such abstractions can be designed to be highly practical and low-overhead, requiring only small additions to existing abstractions.

**Bio:**

*Nandita Vijaykumar is a Ph.D. candidate at Carnegie Mellon University, advised by Prof. Onur Mutlu and Prof. Phillip Gibbons. She is also currently a visiting student at ETH Zurich. Her research focuses on the interaction between programming models, system software, and hardware architecture in an increasingly diverse compute landscape, with a focus on memory systems and modern accelerators like GPUs. She is excited about enabling cross-layer full-stack solutions to make future systems highly efficient and easy-to-program. She is the recipient of the Benjamin Garver Lamme/Westinghouse Fellowship at CMU. During her Ph.D., she has been fortunate to intern at Microsoft Research, Nvidia Research, and Intel Labs.*

**Tuesday, April 9, 2019**  
**2:30 pm**  
**JCL 390**  
**Host: Shan Lu**