Dimitrios Skarlatos
University of Illinois at Urbana Champaign

Title: Enabling Microarchitectural Replay Attacks

Abstract:
Recently, computing systems have been under a barrage of microarchitectural side channel attacks. Such attacks allow adversaries to learn a victim process' secrets by monitoring how that victim uses system and hardware resources. Yet, not all hope is lost. The inherent noisy nature of side channels attacks requires the adversary to run the victim many, and potentially hundreds of, times to reliably exfiltrate secrets. This is good news for defenders. It is reasonable to expect that many outsourced applications will only be run once per input.

In this talk I will introduce Microarchitectural Replay Attacks, a novel class of processor vulnerability based around the premise that on modern out-of-order machines, a single logical dynamic instruction may execute multiple times. This can be leveraged by attackers to create new privacy- and integrity- breaking attacks. Finally, I will present a proof-of-concept microarchitectural replay attack in the SGX setting that denoises a notoriously noisy microarchitecural side channel—port contention—in a single run of the victim application.

Bio:
Dimitrios Skarlatos is a PhD student at the University of Illinois at Urbana-Champaign (UIUC), working with Professor Josep Torrellas. His research lies at the intersection of computer architecture and operating systems. He particularly enjoys building practical solutions that improve the performance and bolster (or sometimes break) the security guarantees of computing systems.

Dimitrios is a UIUC College of Engineering Mavis Future Faculty Fellow, the recipient of the W. J. Poppelbaum Memorial Award, the David J. Kuck Outstanding MS thesis award, the UIUC Computer Science Excellence fellowship, and a MICRO TopPicks in Computer Architecture Honorable Mention. He has earned an MS from UIUC and a BS in Electronic and Computer Engineering from the Technical University of Crete in Greece.

Rui Zhang
University of North Carolina

Title: Automated Security Validation for Processor Designs

Abstract:
The recent rise of hardware attacks, such as Spectre and Meltdown, have shown how vulnerable modern computer hardware is. Attacks at the hardware level can compromise an entire system even with secure software. As hardware is often difficult to patch, validating the security of hardware designs at the design time is important. The current state of the art to find vulnerabilities in hardware designs uses assertion based verification. While powerful, the bottlenecks of it lie in how to systematically identify the security critical properties needed for the verification, and how to determine the root cause for any found violations.

In this talk, I will present our work towards automated security validation for processor designs. It consists of two directions: security property identification and exploit program generation. In the first direction, I will first describe a semi-automatic approach, SCIFinder [ASPLOS'17], that identifies security properties of processor designs. It relies on security errata and machine learning techniques to identify security properties from a set of processor invariants. The security properties built can be useful for similar designs, but repeating the whole process for each new design can be tedious. Thus, I will then describe a tool, Transys [S&P'20], that automatically and efficiently translates security properties from one design to similar designs. It uses static program analysis techniques for the translation, and further reduces the effort in developing security properties. In the second direction, I will talk about an end-to-end tool, Coppelia [MICRO'18], that automatically generates exploit programs. The core of Coppelia is a novel hardware-oriented backward symbolic execution engine. It helps designers better analyze, understand, and assess the security threat for vulnerabilities in processor designs.

Bio:
Rui Zhang is a PhD candidate at University of North Carolina at Chapel Hill, advised by Professor Cynthia Sturton. Her research interests lie at the intersection of formal analysis techniques and hardware security. Her work has resulted in formal, automated systems and tools for detecting vulnerabilities and validating the security of hardware designs. Her research has been recognized with a best paper award nomination at MICRO 2018. She was nominated as one of two candidates for Google PhD fellowship representing UNC Chapel Hill in 2016. She received her master's degree from Columbia University in 2015 and her bachelor's degree from Peking University in 2013.

Yatin Manerkar
Princeton

Title: Progressive Automated Formal Verification

Abstract:
Modern computing systems are deeply heterogeneous and increasingly complex, making it challenging for designers and implementers to ensure their correctness. In current architectural design approaches, verification of correctness is often conducted late in the development timeline, post-implementation. However, late-stage verification results in bugs being found and fixed late in development, when the opportunities for redesign may be limited. My research proposes **progressive** automated formal verification, or automated formal verification throughout the system development process. Under this philosophy, a formal model of a system is created and verified at early-stage design, catching design bugs **before** implementation commences. This model can then evolve as the design progresses, and the eventual implementation can be verified against such a detailed formal model to help ensure correctness. The automation of the tools makes them easy to use, while the formal nature of the verification provides strong correctness guarantees.

My dissertation work has enabled progressive automated formal verification of Memory Consistency Model (MCM) properties for parallel architectures. MCMs specify the ordering and visibility rules for memory operations in parallel programs, so MCM verification of implementations is critical to parallel system correctness. I will discuss methodologies and tools for MCM verification of early-stage microarchitectural designs (PipeProof), detailed microarchitectural designs (RealityCheck), and Verilog RTL implementations (RTLCheck). I will conclude with future work plans, which include applying progressive automated formal verification to distributed systems and cyber-physical systems. I also intend to improve hardware design and verification by adapting hardware description languages and high-level synthesis to be more amenable to formal analysis.

Bio:
Yatin Manerkar is a final-year PhD candidate in the Princeton Computer Science department, advised by Prof. Margaret Martonosi. He holds a BASc in Computer Engineering from the University of Waterloo and an M.S. in Computer Science and Engineering from the University of Michigan. He also worked full-time at Qualcomm Research for one year. Yatin's research develops automated formal methodologies and tools for the design and verification of computing systems. His work has led to the discovery of bugs in a lazy coherence protocol, a "proven-correct" compiler mapping for C/C++11 atomics, a commercial compiler, and an open-source processor. He has also contributed to the development of the RISC-V ISA's memory model by finding deficiencies in its draft specification. Yatin's research has been recognised with two best paper nominations, and three of his papers have been honoured for their high potential impact as either Top Picks or Honorable Mentions in IEEE Micro's annual "Top Picks" issue. Yatin is a recipient of the Wallace Memorial Fellowship, one of Princeton's highest graduate honours awarded to approximately 25 PhD students annually for a senior year of their doctoral studies. He also received the 2019 Award for Excellence from Princeton's School of Engineering and Applied Science.

Mohammad Alian
University of Illinois at Urbana Champaign

Title: A Cross-Layer, Hardware-Software Approach Towards Architecting High-Performance Datacenters

Abstract:
Processing the ever-increasing data generated by smartphones, IoTs, autonomous cars, etc. using deep neural networks necessitates an enormous amount of computation, which only can be achieved through distributed and specialized computing. Moreover, industry is moving from traditional processor-centric servers into composable and software-defined infrastructures, where a server rack is defined by fluid pools of compute, memory, and storage that are networked together. The key enabler of efficient distributed computing and composable architectures is a low-latency, high-bandwidth network. In this talk I am going to talk about my recent research works on distributed computing and datacenter network. My talk is divided into two parts. In the first part I will talk about the concept of memory channel networking (MCN) in which I introduce a near-memory processing framework that scales the compute power, memory capacity, and memory bandwidth of a leaf-node server without compromising the application readiness. In the second part, I will explain a near-memory network interface architecture for ultra-low latency networking without compromising the network bandwidth. These works plot a roadmap for designing a high-performance and energy-efficient near-memory server that can seamlessly accelerate the existing datacenter applications.

Bio:
Mohammad Alian is a 5th year PhD candidate at the University of Illinois Urbana Champaign working with Prof. Nam Sung Kim. His research interests are near-memory processing, in-network processing, data-center network architecture, and unconventional computing paradigms. He has 13 published papers in top tier computer architecture and systems conferences/journals and three patent applications. His research has been recognized by one honorable mention in IEEE MICRO top picks and four best paper candidates including one from IEEE/ACM International Symposium for Microarchitecture and one from IEEE International Symposium for High-Performance Computer Architecture. Mohammad is graduating in December 2019.

Radha Venkatagiri
University of Illinois at Urbana Champaign

Title: Democratizing Error-Efficient Computing via Principled Application-Level Error Analysis

Abstract:
The end of conventional technology scaling has led to notable interest in techniques that aim to improve overall efficiency by allowing the system to make controlled errors. We refer to such systems as being error-efficient: they only prevent as many errors as they need to. Some examples of error-efficiency techniques are approximate computing, that trade off output accuracy for performance/energy or low cost (but less-than-perfect) resiliency solutions that let some hardware errors escape as user-acceptable output corruptions. Error-efficient computing paradigms have the potential to significantly change the way we design hardware and software. However, current solutions rely on programmer expertise to provide application-level error specifications which severely limits their scope as such expertise is sparse and can take years to develop for new domains.

My research aims to advance the state-of-the-art in error-efficient computing by developing automatic application-level error analysis tools that mitigate the need for such hard-to-find programmer expertise. Using a hybrid approach of program analysis and (relatively) few error injections, these tools can comprehensively analyze billions of possible errors that can impact a program's execution at low cost. The resulting comprehensive application error profiles can be utilized to build optimal error-efficiency solutions tailored for individual applications. Specifically, my contributions are: (1) A suite of automated application-level error analysis tools (Approxilyzer, gem5-Approxilyzer and Minnow) that can accurately quantify the impact of all errors (for a given error model) in a program's computation and data on its output quality, (2) A framework called Minotaur that shows a principled application of software testing methodologies to significantly improve the speed, accuracy and scalability of automated error analysis techniques over multiple workloads and inputs, and (3) Proof-of-concept workflows that demonstrate the potential of the automatically generated application error profiles to enable custom error-efficiency solutions that meet user requirements. Together, the solutions proposed by my work can democratize the reach of error-efficient computing to both novice programmers as well as emerging application domains.

Bio:
Radha is a doctoral candidate in Computer Science at the University of Illinois at Urbana-Champaign. Her research interests lie, broadly, in the area of Computer Architecture with a specific focus on Approximate Computing, Hardware/Software Resiliency, Error-Efficient Computing and Software Testing. Radha's dissertation work aims to formalize techniques and build tools that enable reliable, low-cost and efficient computing by allowing controlled errors in the system. She was among 20 people invited to participate in an exploratory workshop on error-efficient computing systems initiated by the Swiss National Science Foundation and is one of the select young researchers worldwide to attend the 2018 Heidelberg Laureate Forum. She has also been selected for the Rising Stars in EECS workshop for the year 2019.

Before joining the University of Illinois, Radha was a CPU/Silicon validation engineer at Intel where her work won a divisional award for key contributions in validating new industry standard CPU features. Prior to that, she worked briefly at Qualcomm on architectural verification of the Snapdragon processor. Radha has a Master's in Computer Engineering from North Carolina State University and a Bachelor's in Electrical Engineering from the University of Madras in India.

Yipeng Huang
Columbia / Princeton

Title: Emerging Architectures for Humanity's Grand Challenges

Abstract:
As we enter the post-Moore's law era of computer architecture, researchers are turning to new models of computation to address humanity's Grand Challenges. These new models of computation include analog computing and quantum computing. At a high level, they offer fundamentally different capabilities compared to classical, digital computing machines. These capabilities include simulating natural phenomena using physics as a direct model. The urgent challenges in harnessing these promising models of computation are in connecting them to conventional architectures, and in helping programmers correctly use them.

First, I will talk about how analog accelerators can play a role modern architectures to aid in modeling stochastic and nonlinear phenomena. The key feature of the analog model of computation is that variables evolve continuously, thereby avoiding many problems associated with the step-by-step updating of variables in all digital machines. Using prototype analog accelerators that I helped build at Columbia University, I demonstrate using analog approximate solutions as good initial seeds for GPUs solving scientific computation problems.

Second, I will talk about helping programmers reason about quantum algorithms with the aid of classical inference tools. My research now is in bridging the architectural gap between the theoretical discipline of imagining quantum computer problems and algorithms, and the experimental discipline of building quantum computer hardware. A vital task in bridging that gap is improving researchers' ability to simulate and debug quantum algorithms. I turn to classical inference techniques as a way to contend with high-dimensional data encountered in quantum computing. The novel programming abstraction offers a way to speed up simulating and debugging important quantum algorithms.

Bio:
Yipeng is a postdoctoral research associate at Princeton University, working with Prof. Margaret Martonosi. He received his PhD in computer science in May 2018 from Columbia University, working with Prof. Simha Sethumadhavan. His research interest is in building, programming, and in identifying applications for emerging architectures. These include quantum and analog computer architectures that may uniquely address urgent challenges in scientific computing, but would require new programming tools and architectural abstractions to harness these unconventional models of computation. His dissertation "Hybrid Analog-Digital Co-Processing for Scientific Computation" was nominated by Columbia University Computer Science for the ACM doctoral dissertation award. His work has received attention among computer architecture researchers, who named his papers one of twelve Top Picks among conference papers in 2016, and again as an honorable mention in 2017. He has also received support for his research work through DARPA, in the form of a Small Business Technology Transfer grant to investigate commercial applications.

Gokul Ravi
University of Wisconsin-Madison

Title: Vertical Integration in Computing Systems: Demystifying Hardware's Clock Abstraction

Abstract:
Historically, long-established and deep-set layers of abstractions, across both hardware and software, allowed designers to cope with the complexity of designing sophisticated computer systems. Unfortunately, these abstractions have limited cross-stack innovation across these systems and this lack of vertical integration has left significant untapped opportunity on the table. Thus, there is a need for vertical integration for a new renaissance in computer architecture.

In this talk, I will discuss how my research takes a "vertically integrated" outlook to innovating across the computing system. In varying detail, I will describe how my research contributions exploit the clocking system's structural, functional and differential characteristics and leverage resulting opportunities in several ways. First, in REDSOC, I speed up the out-of-order core via fine-grained data-aware sequence acceleration. Second, in SHASTA I target synergic, fine-grained, spatio-temporally diverse approximate computing via hardware-software co-design. Third, in TNT, I design a modular on-chip network with potential for near-ideal traversal latency. And fourth, in CHARSTAR, I propose an ML-assisted resource allocation and hardware reconfiguration scheme with novel circuit awareness. Finally, I will briefly discuss ongoing work and future directions on more vertically (and horizontally) integrated system design, assisted by intelligent machine learning techniques, targeting classical computing systems as well as rapidly growing non-classical domains.

Bio:
Gokul Subramanian Ravi is a Computer Architecture Ph.D. candidate in the ECE Department at the University of Wisconsin - Madison, advised by Prof. Mikko Lipasti. His research takes a "vertically integrated" outlook to innovating across the the computing system, by loosening traditional hardware abstractions and leveraging them via system-wide optimizations. Specifically, his dissertation research demystifies hardware's abstraction of the clock, exploiting the clocking system's characteristics across a variety of computing layers and substrates, while targeting multiple application domains. His research has appeared at the premier venues for computer architecture and low power electronics and has initiated both industry and university collaborations. Previously, he received his undergraduate degree from BITS Pilani in India (2012), worked as an ASIC Design Engineer at NVIDIA (2012-2013) and completed internships at Qualcomm (2014) and AMD Research (2016).