Statement of Research Interests

Seetal Potluri

I develop secure and reliable computer systems to address hardware vulnerabilities and faults in both cloud and edge computing environments. To that end, my research interests lie at the intersection of *security*, *reliability*, *computer architecture,* and *machine learning (ML)*.

High performance and energy efficiency are both critical to today's hardware systems deployed both at the cloud and the edge. However, performance/energy optimizations and compactness typically come at the expense of increased hardware vulnerabilities and safety concerns. I am focused on developing secure and reliable hardware accelerators while ensuring the added features come with formal guarantees and at the expense of little overheads. My research involves mathematical analysis of the underlying challenges, deploying ML to address them, and scalable hardware implementations. I ultimately aim to design metrics for evaluating security/safety under performance/energy constraints, and tools that can provide guarantees and enable automated trade-offs between the desired level of security, reliability, performance, and cost.

Proposed Research Plan

I plan to pursue a research program on hardware-based security and reliability consisting of several research thrusts tackling different but related aspects of hardware-targeted attacks and safety concerns. Specifically, I intend to work on trustworthy ML systems, scalable and reliable FPGA virtualization, and automated reverse engineering to detect FPGA malware with a future focus on explainable ML for hardware security. This research will build on a body of work I developed during my Ph.D. and broad post-doctoral and industry research experience.

1. Trustworthy ML Systems

How to secure ML hardware/software systems against model reverse engineering attacks?

Advanced ML classifiers have drastically improved the ability of decision-making. The ability to reverse engineer (RE) an ML model is a serious violation of intellectual property and potentially the confidential data used for training. Unlike cryptography, security was not originally considered for this class of systems but has become an important topic of active research due to the adversarial demonstrations published and demonstrated during the last decade. Rectified linear unit (ReLU), due to its non-bijective nature, by definition, also comes with the additional benefit of better RE resilience compared to its bijective counterparts like *sigmoid*. However, it was recently shown that it is possible to perform RE for both the *architecture as well as the model parameters* of a ReLU network of up to two layers by querying it [1].

There are two hardware accelerator spaces where ReLU networks are actively deployed. For ML edge accelerators deploying deep ReLU networks, I show that using hardware internal states, it is possible to perform *model RE* using *linear constraint satisfaction* with multiple orders of magnitude fewer queries, and better accuracy than through application programmer interface (API) [2, 3]. Since the hardness of constraint solving is a strong function of weights and biases, I am interested to exploit this observation through *resilient training*. One of the initial goals is to train the model by exploiting the limited input numerical precision such that the simplex shrinks iteratively during backpropagation and eventually becomes a null set for as many neurons as possible, thereby ensuring a significant drop in RE accuracy.

The second space is the cloud where the user has access to the ML prediction API. For deep ReLU networks defining piecewise linear functions, it was shown that selective neuron zeroing and measuring the boundary intersections between the linear regions allow for model RE up to isomorphism [1]. I plan to defend this adversary, by reducing the intersections between bent hyperplanes and pushing towards violating linear regions assumption through resilient training with little impact on the classifier's accuracy. Additionally, existing works can empirically demonstrate the complete RE of only up to practical 2-deep ReLU networks. Since real-world networks are much deeper, my focus is to develop more advanced methods to approximate the bending phenomenon and explore RE, while providing mathematical guarantees on the security of >2 deep networks if any. Also, existing model RE works are mostly restricted to fully connected networks, and I am interested in diversifying into convolutional neural networks.

My experience in security of ML accelerators and design-for-test [3-13] puts me in a unique position to conduct this research. I have submitted a Semiconductor Research Corporation (SRC) proposal on this topic, which was successful in phase-I. Although the full proposal was not successful, I plan to incorporate the feedback and resubmit once I join. I am planning short-term research on secure ML accelerators and in the long run, my goal is to provide end-to-end security in ML systems. I already have active collaborations in this thrust area who work on theoretical or system aspects, but I am also looking forward to building further partnerships with researchers working in theory. I am planning to write a proposal to the National Science Foundation (NSF) Secure and Trustworthy Cyberspace (SaTC) program under the 'CORE' designation and 'small' class, since they emphasize strong mathematical foundations, principled design methodologies, and metrics for success/failure in cybersecurity. SRC hardware security (HWS) is another suitable program to sponsor this line of research.

2. Scalable and Reliable FPGA Virtualization

How to scale FPGA virtualization and improve partial reconfiguration speed, while ensuring reliability?

Clouds leverage virtual machines (VMs) to allow computing as a pay-per-use utility with market-oriented allocation policies. Owing to their flexibility and superior performance, cloud providers nowadays offer on-demand FPGA instances for application acceleration. Due to the massive parallelism available in these instances e.g., Amazon's EC2 F1, supporting multiple VMs simultaneously on a single FPGA, known as *spatial sharing*, improves resource utilization. To support cases when multiple VMs do not fit on the FPGA, vendors like Xilinx also support the partial reconfiguration (PR) feature, which although incurs significant performance overhead, can facilitate *temporal sharing* of resources across VMs. However, system support is still in its infancy and researchers are systematically working on building the system software frameworks and more recently on security, fairness, energy efficiency, etc.

Based on the literature and my experience at Xilinx and interactions with Amazon, there are two important issues blocking the active deployment of FPGA virtualization. First is scalability: current works are point solutions, thereby with each new feature/observation, the designers are forced to completely redesign the system software, which is laborious, expensive, impractical, and hence not scalable. To address this, I plan to exploit ML to track different inefficiencies in the prior cycles e.g., idle times, context switches, fault profiles, etc. to make sure they are compensated for in the long term. In the past, such methodologies have been successfully deployed to improve the quality of service for radio access networks by VMWare. The advantage of this framework is that if future FPGAs present any new characteristics or issues, it is sufficient to add new constraints in the training function and not necessary to redesign the system software, thus making the approach *scalable*. To realize the vision of building a scalable FPGA virtualization, I am looking forward to collaborating with researchers in the department (or across the departments) working on operating systems, and computer architecture.

The second issue is partial reconfiguration: the PR time is a major bottleneck for deploying FPGA virtualization. Researchers have proposed several solutions, yet they cannot reconfigure within 10,000 clock cycles even for relatively small partitions. We have recently shown that the excessive peak power dissipation during PR of a clock region while running power-hungry workloads in other clock regions, causes PR failures and in extreme cases device shutdown [14]. Hence, while increasing the reconfiguration speed, we also have a contrasting requirement for IR-drop. Similar observations were recently made by Meta and Google on silent data corruption errors in their data center workloads, mainly due to computationally intense ML applications.

I am one of the few researchers in the US actively working on reliability issues during partial reconfiguration, and planning to extend my expertise in scan IR-drop to realize FPGA acceleration. I intend to write my NSF CAREER proposal on "FPGA Virtualization: Efficiency, Reliability, and Security" for a target timeline of 2023-2027. This proposal will include my prior work on reliable FPGA reconfiguration and ongoing work on fair allocation and energy efficient scheduling. I am also planning to develop a course to teach FPGA virtualization to the next generation of engineers and potentially use this experience in my NSF CAREER proposal. Once I have the initial results, I also plan to talk to my contacts at Xilinx Research to attract further funding on this topic.

I assisted my post-doc supervisor on proposal preparation for the Defense Advanced Research Projects Agency (DARPA) call for run-time reconfigurable arrays in less than 100 clock cycles, and more recently on the Processor Reconfiguration for Wideband Sensor Systems (PROWESS) for reconfiguration within 50ns for RF applications. Similarly, because my research on FPGA virtualization at NC State is funded by the ONR, I know their next developments and specific interests on this topic as well. Due to the increased interest in multiple funding agencies, I am looking for collaborations with experts in the RF domain, to write joint proposals on this topic. Research on this thrust would require electronic design automation (EDA) tools for design, simulation, chip tape-out, printed circuit board design for characterization, as well as measurement equipment like high-end oscilloscopes and pattern generators.

3. Reverse Engineering for Malware Detection in Cloud FPGAs

How to reverse engineer client designs for malware using formal methods, while maintaining privacy?

Active deployment of FPGAs for domain-specific customization in data centers has raised multiple concerns due to malware that provides user access to privileged data, or stealthy faulty behaviors of user logic [15]. Examples of these attacks include denial-of-service [14], adversarial hijacking, remote misclassification, and cross-VM covert channels. While formal methods have been actively applied to provide security guarantees for secure access control, there are no detection solutions with formal guarantees for malware. The user gets access to the cloud resource by first submitting his/her register transfer level (RTL)/C code to the cloud provider, where an application engineer (AE) manually checks the code, before granting the resource. The user subsequently uses vendor-specific EDA tools to synthesize, implement, and finally generate a bitstream targeting specific physical regions on the FPGA.

I have worked earlier on RE attacks and formally guaranteed defenses for both digital circuits [7, 8, 16] and biochips [17, 18]. I am interested in diversifying into RE in the context of FPGAs from an attack detection perspective. The manual checking of the RTL/C by the AEs in the cloud is laborious, error-prone, and hence not scalable. I plan to automate this by converting to an intermediate representation and subsequently use equivalence checking tools to verify against a predefined set of properties. Since formal verification is in general time-consuming, I am also interested in exploring efficient data structures that exploit specific properties and provide quick solutions/answers. On most occasions, for privacy reasons, the user is not interested in sharing the RTL/C with the cloud provider, I am interested in exploring secure two-party RTL/C verification frameworks to address such prevalent scenarios.

The second challenge of the cloud provider is to detect malware inside the bitstreams once they reach the clock region. I plan to use formal methods for bitstream authentication with guarantees. While there has been recent work that uses ML to detect malicious circuits with few false positives/negatives [19], it is restricted to ring oscillator (RO) circuits and needs to be generalized to all workloads, provide formal guarantees, and extend to partial bitstreams. Additionally, it is assumed that the FPGA owner has access to the encryption key, thus being able to decrypt the bitstream before performing RE, which may not be acceptable to a benign user. I plan to address these issues by proposing new sensitivity metrics that formally quantify the impact of the malware on different portions of the system and propose detection methods using these metrics with strong guarantees. Finally, to address the benign user challenge, I plan to establish a secure two-party or homomorphic encryption framework to protect the user's privacy while enabling the cloud provider to verify designs prior to reconfiguration.

I would like to complement my expertise in reverse engineering, FPGAs, constraint solving, with collaboration of experts in formal methods and theoretical cryptographers to solve these challenges. I plan to reuse the infrastructure built for the second thrust. My plan is to first establish a credible research group on the first two thrusts and then apply for industry grants on secure EDA for a target timeline of 2025-2028.

4. Explainable ML for Hardware Security

Is it possible to explain why ML improves the attack efficacy, to assist robust defense design?

There is a significant body of work in recent years that shows the efficacy of ML in breaking defenses. While this is positive in the sense that one could use this information to build more robust defenses, this kind of information is often difficult to obtain from ML classifiers. For example, we have recently demonstrated the effectiveness of 2D deep learning to break parallel implementations of post-quantum key exchange protocols using Gramian angular sum fields (GASF) [20]. While we know that GASF exploits the temporal correlation of time-series data thereby playing a part in image classification, the exploited weakness of the implementation is not revealed, thereby not providing adequate information enough to build a robust defense. Similarly, we have recently shown that ML is able to successfully detect randomly injected dummy states and jitter-based hiding countermeasures for quantum-resistant hardware [21, 22], however it is not clear how the classifier is able to eliminate the effects of the defenses, making robust defense design challenging. Orthogonally, there have been some successful attempts in using neural networks trained using hardware performance counters for real-time intrusion detection. However, the lack of proper explanation about why specific sets of microarchitectural features correlate well to specific traits of malware remains a critical bottleneck. The main challenge in providing explanations is having domain expertise in two different topics, one being ML and the other dependent on the underlying hardware security problem. This research requires the collaboration of a hardware security expert like me and the other being a theory expert in ML.

Funding Opportunities

In addition to the funding programs that I discuss in my research plans such as NSF CAREER and SaTC, DARPA, ONR, and Xilinx, there are additional opportunities to sponsor my research. The development of FPGA virtualization is of commercial interest to cloud providers like Amazon and Microsoft. There is also increased interest in expeditionary and intelligent microsystems – self-tuning, self-optimizing, and mission reconfigurable, at an acceptable size, weight, power, and cost (SWaP-C). These systems are expected to typically run mission-critical and possibly safety-critical workloads and require a deterministic level of performance. Therefore, research on FPGA virtualization, security, and reliability is suitable for Sandia National Laboratories, Air Force Research Lab (AFRL), the National Institute of Standards and Technology (NIST), the United States Military Academy (USMA), and the Air Force Office of Scientific Research (AFOSR) proposals. Privacy issues in the cloud are a threat to applications processing sensitive healthcare data, hence this research is also suitable for National Institute for Health (NIH) proposals. I have also been in close contact with researchers working on formal methods for hardware security verification at Intel Offensive Security Research labs and Semiconductor Research Corporation (SRC) industry liaisons and researchers in charge of cybersecurity funding programs at Google, Meta, IBM, and ARM.

References

(1) David Rolnick, and Konrad P. Körding, "Reverse Engineering Deep ReLU Networks", *International Conference on Machine Learning (ICML)*, 2020, pp. 8178-8187.

(2) **S. Potluri,** and A. Aysu, "Stealing Neural Network Models through the Scan-Chain: A New Threat for ML Hardware", *IEEE International Conference on Computer Aided Design (ICCAD)*, 2021, pp. 1-8.

(3) **S. Potluri**, et al., "LPScan: An algorithm for supply scaling and switching activity minimization during test", *IEEE International Conference on Computer Design (ICCD)*, 2013, pp. 463-466.

(4) **S. Potluri**, et al., "DFT Assisted Techniques for Peak Launch-to-Capture Power Reduction during Launch-On-Shift At-Speed Testing," *ACM Transactions on Design Automation of Electronic Systems (TODAES)*, 2015, Vol. 21, No. 1, pp. 14:1-14:25.

(5) **S. Potluri**, N. Chandrachoodan, and V. Kamakoti, "Interconnect Aware Test Power Reduction", *Journal of Low Power* Electronics, Vol. 8, No. 4, 2012, pp. 516-525.

(6) **S. Potluri**, A. Aysu, and A. Kumar, "SeqL: Secure Scan-Locking for IP Protection", *IEEE International Symposium on Quality Electronic Design (ISQED)*, 2020, pp. 7-13.

(7) **S. Potluri** et al., "SeqL+: Secure Scan-Obfuscation with Theoretical and Empirical Validation", IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems (TCAD), 2022 (Accepted).

(8) **S. Potluri** et al., "Cell-Aware ATPG to Improve Defect Coverage for FPGA IPs and Next Generation Zynq MPSoCs," *IEEE Asian Test Symposium (ATS)*, 2017, pp. 157-162.

(9) **S. Potluri**, et al., "PinPoint: An algorithm for enhancing diagnostic resolution using capture cycle power information", *IEEE European Test Symposium (ETS)*, 2013, pp. 1-1.

(10) **S. Potluri**, A. S. Trinadh, S. Saraf and K. Veezhinathan, "Component fault localization using switching current measurements," *IEEE European Test Symposium (ETS)*, 2016, pp. 1-2.

(11) **S. Potluri**, P. Pop and J. Madsen, "Design-for-Testability of On-Chip Control in mVLSI Biochips," in *IEEE Design & Test*, vol. 36, no. 1, pp. 48-56.

(12) G. Haas, **S. Potluri** and A. Aysu, "iTimed: Cache Attacks on Apple A10 Fusion SoC," *IEEE International Symposium on Hardware Oriented Security and Trust (HOST)*, 2021, pp. 80-90.

(13) E. Karabulut, C. Yuvarajappa, M. I. Shaikh, **S. Potluri**, A. Awad, and A. Aysu, "PR Crisis: Analyzing and Fixing Partial Reconfiguration in Multi-Tenant Cloud FPGAs", *ACM Workshop on Attacks and Solutions in Hardware Security (ASHES),* 2022 (Accepted).

(14) S. Trimberger, and S. McNeil, "Security of FPGAs in data centers," *IEEE International Verification and Security Workshop (IVSW)*, 2017, pp. 117-122.

(15) Q. Tan, **S. Potluri,** and A. Aysu, "Efficacy of Satisfiability-based Attacks in the Presence of Circuit Reverse-Engineering Errors", *IEEE International Symposium on Circuits and Systems (ISCAS)*, 2020, pp. 1-5.

(16) H. Chen, **S. Potluri,** and F. Koushanfar, "BioChipWork: Reverse Engineering of Microfluidic Biochips", *IEEE International Conference on Computer Design (ICCD)*, 2017, pp. 9-16.

(17) H. Chen, **S. Potluri,** and F. Koushanfar, "Security of Microfluidic Biochip: Practical Attacks and Countermeasures", *ACM Transactions on Design Automation of Electronic Systems (TODAES)*, Vol. 25, No. 3, Art. 27, 2017.

(18) R. Elnaggar, et al., "Learning Malicious Circuits in FPGA Bitstreams," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems (TCAD)*, 2022.

(19) P. Kashyap, F. Aydin, **S. Potluri**, P. Franzon, and A. Aysu, "2Deep: Enhancing Side-Channel Attacks on Lattice-Based Key-Exchange via 2D Deep Learning", *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems (TCAD)*, Vol. 40, No. 6, 2021, pp. 1217-1229.

(20) F. Aydin, P. Kashyap, **S. Potluri,** P. Franzon, and A. Aysu, "Breaking Side-Channel Countermeasures through Deep Learning", *IEEE International Conference on Computer Aided Design (ICCAD)*, 2020 (Invited).

(21) F. Aydin, **S. Potluri**, and A. Aysu, "Machine Learning for Side-Channel Assessment of Next Generation Cryptosystems", *IEEE International Symposium on On-Line Testing and Robust System Design (IOLTS)*, 2022 (Invited).