Guest Editors' Introduction to the Special Issue on Hardware Security

Amro Awad, *Member, IEEE* and Rujia Wang, *Member, IEEE*

HARDWARE security is becoming a significant challenge in modern computing systems. Recently discovered hardware vulnerabilities, such as Spectre and Meltdown, are striking evidence that today's computing systems are untenable without deliberate consideration of the security aspects at the design time. Therefore, this special issue of IEEE Transactions on Computers (TC) focused on hardware security and is particularly interested in making future computing systems more secure and tenable without significant performance degradation. This special issue on hardware security includes 12 outstanding papers on various topics related to hardware security: secure-by-design architectures, secure speculative execution, secure system integration of untrusted chiplets, malware detection, program analysis using power side-channels, architecture support for forensics, and efficient implementations of security modules.

The articles accepted in this special issue devise new security mechanisms and architectures that can be leveraged in future systems. From new architectures that consider security at the design time, architectural changes to protect against architecture state attacks, efficient mechanisms for detecting malware and various types of attacks, fast and energy-efficient implementations of cryptographic algorithms in hardware security modules, to architectural support for forensics in IoT workloads, we believe this special issue presents a unique set of exciting and timely papers that address a set of the most pressing security challenges in modern processors. We can categorize the papers accepted in the following categories: architecture security, hardware implementation of security primitives, attack detection mechanisms, and forensics.

We have received 53 submissions from all over the world, and only 12 papers were accepted, i.e., the acceptance rate of 22.6 percent. While rare submissions were somewhat of scope for this issue, the majority of submissions were addressing issues relate to hardware security, trusted execution environments, hardware-assisted security support, and secure integration of hardware chips. We treated all papers similarly with regard to the selection criteria, and our decisions are majorly based on experts' recommendations. We have occasionally solicited more reviews for papers with inconsistent scores to help us reach a final decision.

In this special issue, there is a strong presence of papers advocating for security-aware architectures and designs. In particular, four papers propose new architectures, architecture support for security or architecture support for forensics. The first paper, titled *OPTIMUS: A Security-Centric Dynamic Hardware Partitioning Scheme for Processors that Prevent Microarchitecture State Attacks*, proposes a novel architecture that allows dynamic partitioning of shared resources while maintaining strong security guarantees and protections against microarchitecture state attacks. OPTIMUS supports deterministic resource allocation at application granularity, which limits the number of hardware reconfigurations allowable to meet the bounded information leakage and resistance against timing and termination channels.

Our special issue also includes a paper *Built-in Security Computer: Deploying Security-first Architecture Using Active Security Processor*. This paper advocates for a new architecture that leverages active security processors and discusses the design trade-offs and optimizations required to realize such designs. The paper implements an off-chip Active Security Processor (ASP) and evaluates its security and performance in real hardware. The ASP runs several security tasks and is completely isolated from the CPU.

One of the accepted papers, titled *Understanding Selective Delay as a Method for Efficient Secure Speculative Execution*, explores using selective delays as a way to defend against speculative execution attacks. The paper investigates the impact of delaying memory loads on memory level parallelism. The paper also reevaluates value prediction as an invisible form of speculation. The paper evaluates the limiting factors of using value prediction in this context and reveals that the accesses required for validation of prediction can potentially introduce a new type of unsafe speculation.

With relevance to vulnerabilities of emerging memory technologies, an accepted paper, *Enabling Secure NVM-based in-Memory Neural Network Computing by Sparse Fast Gradient Encryption*, investigates the security challenges for in-memory neural network computing using emerging non-volatile memories (NVMs). The paper argues that the data remanence issues of emerging NVMs will require protecting the confidentiality of the weights stored in memory and used for in-memory computations. Later, the paper proposes two novel optimizations, one called sparse fast gradient encryption

- *Amro Awad is with the Electrical and Computer Engineering Department at North Carolina State University, Raleigh, NC 27695. E-mail: ajawad@ncsu.edu.*
- *Rujia Wang is with the Computer Science Department at Illinois Intitute of Techonlogy, Chicago, IL. E-mail: rwang67@iit.edu.*

Digital Object Identifier no. 10.1109/TC.2020.3021223

and the other is for scheduling encryption at run-time. The paper shows that such security support will incur marginal overheads. In one accepted paper, *2.5D Root of Trust: Secure System-Level Integration of Untrusted Chiplets*, the authors for the first time discuss how to leverage the 2.5D interposer technology to establish system-level security. The design can allow system vendors to procure chiplets from the open markets while only producing trustworthy inter-poser and assembling the system oneself. The proposed security-enforcing interposer is able to check the system level activities of all untrusted commodity chiplets against security policies via physically separated security features.

There are two accepted papers related to power side-channel attacks. In the paper titled *SCAUL: Power Side-Channel Analysis with Unsupervised Learning*, the authors introduce side-channel analysis with unsupervised learning that can recover the secret key without requiring prior knowledge on FPGA. The technique utilizes an LSTM auto-encoder to extract features from power traces with high mutual information with the data-dependent samples of the measurements. On a lightweight implementation of AES on Artix-7 FPGA, the results show that SCAUL is able to recover the correct key with 3,700 power measurements with random plaintexts, much less than existing DPA based attack.

In another paper, *Instruction Sequence Identification and Disassembly Using Power Supply Side-Channel Analysis*, the authors identify that instruction-level activity could be leaked via a power-based side channel. A dynamic programming algorithm is applied to detect the boundaries of instructions in a sequence with 100 percent recovery rate. A unique aspect of this technique is the use of multiple power supply pin measurements to increase precision and accuracy. With their technique, power leakage data from ten target FPGAs programmed with a prototype of the pipelined architecture was analyzed, and classification accuracy averaging 99 percent were achieved when instructions are grouped based on hardware utilization.

In this special issue, there are two papers target at security issues on IoT systems and architecture. In *MTHAEL: Cross-Architecture IoT Malware Detection Based on Neural Network Advanced Ensemble Learning*, the authors propose a robust cross-architecture IoT malware hunting model based on advanced ensemble learning. The unique MTHAEL model using a stacked ensemble of heterogeneous feature selection algorithms and state-of-the-art neural networks to learn different levels of semantic features demonstrates enhanced IoT malware detection than existing approaches. MTHAEL is the first of its kind that effectively optimizes RNN and CNN with high classification accuracies and consistently low computational overheads on different IoT architectures.

Another paper, *A Hardware-Based Architecture-Neutral Framework for Real-Time IoT Workload Forensics*, proposes a hardware-based generic framework for IoT workload forensics, an infrastructural technique to securely monitor and ensure delivered IoT services in accordance with specifications and regulatory compliance. In particular, this technique identifies digital workloads being executed in real time through dynamic program behavior modeling based on architecture-level data, fulfilled by dedicated machine learning hardware, without high-level software intervention. In contrast to the conventional software-based solutions, whose effectiveness may be undermined by software attacks and introduce significant runtime overhead, a hardware-based framework enables a secure, prompt, and non-intrusive solution.

At last, there are several papers focus on improving algorithm efficiency by considering different hardware properties. We have one accepted paper targets at side-channel analysis of a quantum-resistant altorithm. In the paper *Side-Channel Analysis and Countermeasure Design on ARM-based Quantum-Resistant SIKE*, the authors thoroughly evaluate a post-quantum standardization process called supersingular isogeny key encapsulation (SIKE) with side-channel analysis. This work shows practical and effective attacks can be done with SIKE implemented on real ARM-based devices. Also, an efficient window-based countermeasure is proposed to eliminate the vertical leakage and prevent side-channel attacks with low overhead.

One accepted paper, called *A Lightweight Detection Algorithm For Collision-Optimized Divide-and-Conquer Attacks*, proposes a novel algorithm that exploits correlation-enhanced collision attack to optimize template attacks, and proposes a lightweight collision detection algorithm that enables reducing the search space for keys. The proposed method exploits a jump detection mechanism to efficiently reduce the repetitive collision detection on chains with the same prefix sub-chains. The paper also proposes guessing theory to reorder the collision detection of the sub-keys according to their guessing lengths.

In the paper of *Elliptic Curve Cryptography Point Multiplication Core for Hardware Security Module*, the authors propose a hardware-friendly elliptic curve cryptography implementation that can be utilized widely on different system platforms. The proposed RNS hardware architecture can support fast elliptic curve point-addition (ECPA), point-doubling (ECPD), and point-tripling (ECPT). The authors implemented different ECC point multiplication algorithms on the Xilinx FPGA platform.

As guest editors, we would like to express our sincere gratitude to all authors who submitted their work to this special issue. We also would like to thank all the anonymous reviewers for their valuable time in evaluating and judging the submissions. Further on, we would like to thank the Editor-in-Chief Dr. Ahmed Louri, Associate Editors Dr. Avinash Karanth and DR. James Hoe, Corresponding Topical Editor Dr. Jun Yang, and all the staffs of the IEEE Transactions on Computers for their continuous help and guidance to make this Special Issue possible.

Amro Awad (Member, IEEE) received the PhD degree in computer engineering from NC State University, in 2016. He is currently an assistant professor with North Carolina State University. Before joining academia, he was a senior member of technical staff (SMTS) at Sandia National Laboratories. Moreover, he has been summer faculty fellow with the US Air Force Research Laboratory, in 2018 and 2020. His research is focused on secure hardware architectures and emerging memory technologies.

Rujia Wang (Member, IEEE) received the bachelor's degree from Zhejiang University, and the MS and PhD degree in electrical and computer engineering from the University of Pittsburgh. She is currently an assistant professor of computer science at the Illinois Institute of Technology. Her research interests include the broader computer architecture and systems area, including scalable, secure, reliable, and high-performance memory systems and architectures.