Demo: Analog Trojan Design, Fabrication, and Detection**

Yumin Hou*, Hu He*, Kaveh Shamsi†, Yier Jin†, Dong Wu* and Huaqiang Wu*
*Department of Microelectronics, Tsinghua University
†Departement of Electrical and Comptuer Engineering, University of Florida
yier.jin@ece.ufl.edu
**The hardware demo is based on a HOST 2018 paper titled "R2D2: Runtime Reassurance and Detection of A2 Trojan"

***Abstract*—This demo will show a fabricated analog hardware Trojan detection method. This design mainly targets on A2-like Trojans which are triggered by a successive toggling events. The principle of the detection method is to guard a set of concerned signals, and initiate a hardware interrupt request when abnormal toggling events occur in these guarded signals. We design a processor based on ARMv7-A&R ISA, and insert an analog Trojan into the processor. The detection mechanism is also implemented in this processor. We also fabricate a proof-of-concept chip in the SMIC 130 nm Mixed-Signal 1P7M process and demonstrate that the analog Trojan works on the ARM processor, and the detection method is effective in detecting the analog Trojan.**

I. HARDWARE ARCHITECTURE

A. The ARM-compatible Processor

We propose a fused microarchitecture based on the ARMv7-A&R ISA. This ARM processor is named Merlin. Using michroarchitectural techniques, Merlin expands the DSP capabilities of the ARM processor. Merlin supports most traditional ARM instructions, but does not support some ISA extensions, such as Thumb, ThumbEE, Jazelle, Floating-point, and Advanced SIMD. We realize 181 ARM instructions in total, which is enough to run common benchmarks, such as DhryStone, CoreMark, DSPStone, and EEMBC telecom. There are 7 execution modes defined in ARMv7-A&R ISA, while Merlin works only under user mode. Merlin adopts a fused microarchitecture integrating in-order superscalar and VLIW. Normally, Merlin works under dual-issue in-order superscalar mode. It can be switched to 6-issue VLIW mode when the task is compute intensive. Mode switch can be performed through software. Merlin can be used as an MCU, or a DSP under different application scenarios. Merlin has 16 KB of on-chip L1 instruction Cache, with a 256-bit wide port, and 32 KB dual-port data memory, with each port being 64-bits wide. Merlin has 6 functional units, consisting of 2 arithmetic units (A), 2 multiply units (M), and 2 load/store units (D).

An SoC is designed, where Merlin is used as an MCU. The chip diagram is shown in Figure 1. On this chip, Merlin is integrated with DMA, ROM, SRAM, four 128KB embedded ReRAM, and a variety of peripheral I/O. The Dhrystone performance of Merlin is 1.9 DMIPS/MHz, which is comparable to ARM Cortex-A8 processors.

B. A2-like Analog Trojan in Merlin

A2 is a small analog circuit, which can be inserted into an already placed and routed design. It reads a digital pulse signal (the trigger input), and triggers the payload when the pulse signal has toggled with a high frequently for a certain period of time. The trigger input is connected to a signal that can be toggled with high frequency through a special code snipper running on the processor. The attacker insures that the trigger signal has a much lower toggle rate during typical workloads. This makes detecting the hardware Trojan difficult through testing, not to mention that there exists many low toggling frequency bits in modern processors.

When inserting the analog Trojan trigger circuit into the Merlin processor, we should first select a viable trigger input. The trigger input should have low toggling rate in common cases. It should be controllable through software, so that the trigger code can make it toggle at high frequency to launch the attack.

CPSR (Current Program Status Register) is a software reachable register. The definition of CPSR is shown in Figure 2. We select the CPSR_J bit as the trigger input. Since Merlin does not support ISA extensions, this bit has no function. We use R0 as the attack payload. Once the attack is triggered, the value store in R0 will be modified. We generate the trigger input by frequently writing 0 and 1 to CPSR_J alternatively. When the Trojan is triggered, it changes the value stored in R0 from 0 to 1 so that we can observe the change through a register read. We also attach the trigger output signal

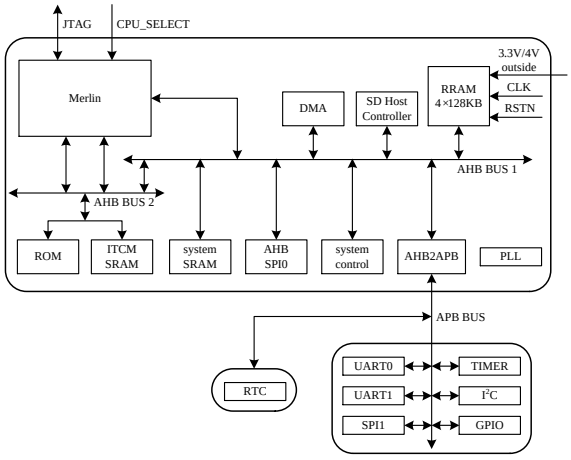

Figure 1: The SoC chip diagram

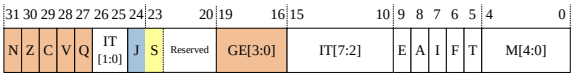

Figure 2: ARM CPSR register

to GPIO, so that we can observe this signal via a oscilloscope.

C. Analog Trojan Detection Scheme

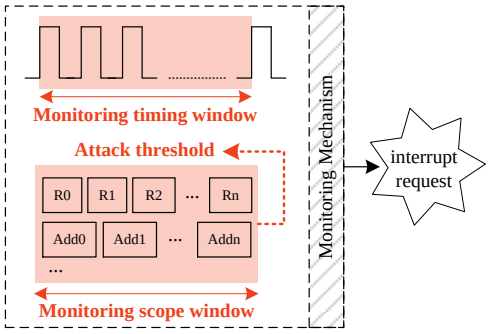

Figure 3: Mechanism of the hardware Trojan detection method

The principle of this method is to guard a set of concerned software controllable registers or memory related signals. A hardware interrupt will be generated if abnormal toggling events occur in the guarded items. The mechanism cannot be disabled through unprivileged software. As shown in Figure 3, R2D2 has several parameters that must be tuned in order to ensure the effectiveness of the scheme and eliminate false positives. The first parameter is the monitoring timing window size. During a monitoring window, the detection unit counts the toggling events on the concerned signal. Throughout this time window, if the toggling frequency increases beyond the attack threshold, the detection circuit will generate an interrupt request. The other important parameter is the monitoring scope which decides the signals to be selected for monitoring. These signals must be ones that have a low toggling frequency during normal processor workloads. Each guarded signal can have a different attack threshold value.

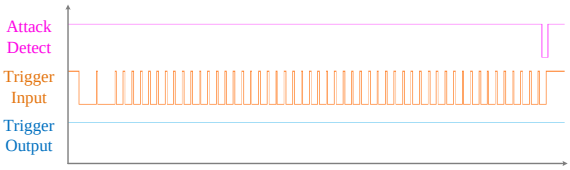

Figure 4: R2D2 detection circuit simulation result

The demonstrated analog Trojan detection method is proved effective through simulation, as shown in Figure 4. The attack_detect signal generates a low level pulse when the toggling events of the guarded signal reaches the attack threshold.

II. Hardware Demo

A PCB board is fabricated, as shown in Figure 5. The Trojan trigger code is stored in a flash. When the PCB board is

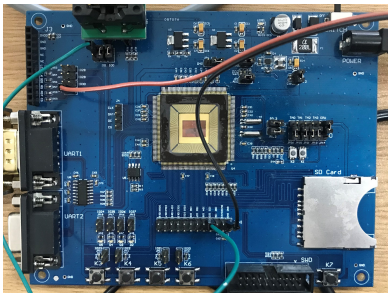

Figure 5: The PCB board

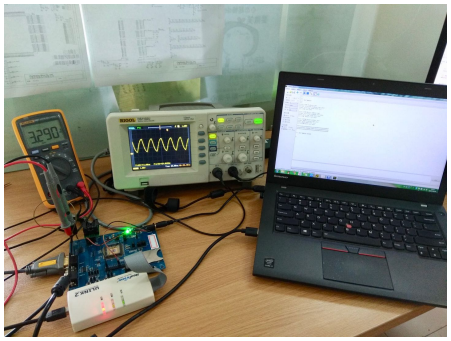

Figure 6: Test environment

power-on, the processor starts executing the program stored in the flash automatically. The Trojan trigger output signal is connected to GPIO. We can observe this signal using a oscilloscope.

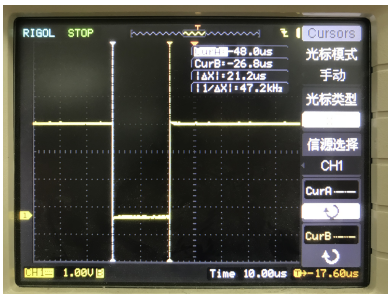

Figure 7: The trigger output signal observed through a oscilloscope

Figure 7 shows the observed trigger output signal. It is a 21.2 us width low level pulse signal. In different experiments, the width of the pulse signal varies in the range between 20 us and 30 us.

III. Conclusion

In this hardware demo, we show an analog Trojan in a in-house designed ARM processor. We will then demonstrate runtime Trojan detection method. The method targets Trojans triggered by toggling events, overcoming a significant limitation of existing Trojan detection schemes in detecting A2-alike Trojans. This method is proved to be effective in detecting an analog Trojan inserted in the ARM processor. The chip is also fabricated using SMIC 130 nm Mixed-Signal 1P7M process.