Inherently Embedded Hardware Trojans

M. R. Strong, K. Oppong Banahene, R. L. Geiger, D. J. Chen
Department of Electrical and Computer Engineering
Iowa State University
Ames, IA, USA

Abstract— **One aspect of system security is evaluating a system's vulnerability to Trojan attack. A hardware Trojan attack can have potentially devastating effects, especially given the increased reliance on integrated circuits within critical systems. A significant amount of research concerns attacks on digital systems, but attacks on AMS and RF systems have recently been of interest as well. A class of Trojans has been proposed that uses undesired alternate modes of operation in nonlinear systems as the Trojan payload. These Trojans are of particular interest because they do not cause deviations from the ideal system performance and cannot be detected until the Trojan is triggered. This work addresses this class of Trojans by listing different payloads, trigger mechanisms, and examples of system architectures vulnerable to attack.**

Keywords—hardware security, hardware Trojan, nonlinear system, PAAST Trojan

I. Introduction

Many electronic designs employ architectures with multiple modes of operation. The existence of multiple solutions may pose a problem for the desired operation of a system. In these situations, engineers include start up circuits or control feedback to prevent undesired operating modes. In other cases, the nonlinearity of the system is essential for its operation. A motivating example is the bistable latch. This system has two stable equilibria, and the system can be forced from one equilibrium to the other through an appropriate input. This begs the question; can a system be driven into an undesired mode of operation? If so, the undesired mode can be viewed as a hardware Trojan payload.

As is the case for any hardware Trojan, a target of opportunity is selected by an adversary. The adversary then designs the system such that a malicious function can be carried out through the Trojan payload, usually after some trigger has activated the Trojan. What makes the proposed class of Trojans so insidious is that the system performance is identical to the desired system performance while not in the Trojan state, and hence the Trojan cannot be detected by measurements or simulations alone.

These hardware Trojans have been previously referred to as Power, Area, Architecture, and Signature Transparent (PAAST) Trojans [1]. In this work we will classify the systems studied in previous PAAST Trojan research as well as related Trojans that can be embedded in other common nonlinear systems. The classes will be defined by the type of payload and examples will be provided for each class.

II. Classification by Payload

A. Autonomous System Classes

For the purpose of clarity, a system will be called autonomous if it does not possess an input as defined by the intended user. The reason for this statement is that an autonomous system's behavior is affected by certain parameters, like the power supply voltage, that are generally considered to be of a constant value and thus not defined as an input by the intended user. However, an adversary may manipulate one of these parameters to trigger the payload. Variations of these parameters are typically referred to as perturbations. Manipulations of these parameters by an adversary will also be called perturbations.

Some circuits have a desired DC output that is expected to be relatively invariant over the operational range of the system. These circuits are typically used as a reference to bias other circuits in the system. If a given output has more than one stable DC solution, the system may undergo a *change in reference*. Several common circuit architectures that can have a change in reference are discussed in [2] and [3].

There are some circuits that are designed as a reference with one or more invariant DC outputs, but these circuits may also oscillate under certain conditions. Conversely, some circuits may be designed to oscillate but can produce a DC output instead under some conditions. In these circuits, we say that the system can have a *change in stability*. Many oscillators have this characteristic and a designer may take steps to ensure proper start up of the system. It was shown in [4] that the Rambus oscillator can have a change in stability for a small range of inverter size ratios.

Another autonomous system class is comprised of oscillators with multiple dynamic modes of operation. A circuit can be designed to have a single oscillatory output with an ideally fixed amplitude, a fundamental frequency, and harmonic content, or the circuit may have multiple oscillatory outputs each with unique amplitude and frequency characteristics as well as a phase relationship with the other outputs. If the circuit output is capable of at least one other distinct periodic solution, the system is said to have a *change in spectrum*. The nonlinear feedback elements in the Wein Bridge oscillator can introduce a

DISTIBUTION STATEMENT A. Approved for public release: distribution unlimited.

change in spectrum where two or more oscillatory outputs differ in amplitude, frequency, or both [5]. Cross-coupled oscillators with multiple oscillatory outputs that differ in phase and frequency are described in [1] and [5].

B. Nonautonomous System Classes

A system with an input as defined by the intended user is called a nonautonomous system. These systems may be triggered into an alternate mode of operation through perturbation like in the autonomous systems, or the payload may be triggered through the input channel.

Controlling feedback can be used in a circuit to regulate the output or force the output to track an input. Stability of the controller must be considered during design. If the controller does not perform the desired function, the system is said to have a *regulation/tracking error*. An example of this is a slew rate enhancement circuit, as shown in [6], that is included in the circuit so the amplifier output tracks with the input.

Some nonautonomous systems have a property called *conditional stability*. The characteristics of these systems that some set of inputs can produce system oscillations that persist even when the input is removed. Active filter architectures with this property were presented in [7].

The final class defined for this work is nonautonomous systems that do not have a unique input-output relationships but do not fall into the other categories. For example, a filter with jump resonance would fall into this category because the spectral characteristics of the output for a given input are not unique. Other circuits in this class are those that have multiple spectral responses for the same input.

III. Trigger Mechanisms

When a Trojan payload is triggered through a model variable that is not a user-defined input, the trigger is called a system perturbation. Perturbation of a system can cause the internal states and the output to approach or coincide with the corresponding states and outputs of an undesired solution.

The power supply voltage in an integrated circuit is typically set by some form of voltage regulator such as an LDO. A perturbation to the power supply voltage can be caused by suddenly increasing the load to the voltage regulator. Similar effects can be caused temporarily by exploiting the bus delay of a supply voltage. The coupled ring oscillator shown in Fig. 1 can be triggered into an undesired mode of oscillation by temporarily mismatching the supply voltage, which is shown in [1].

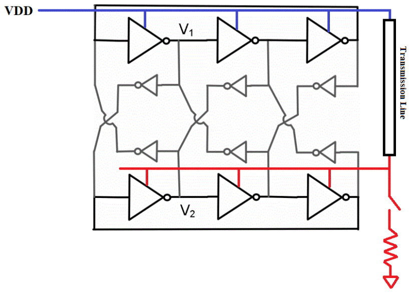

Figure 1: Coupled Ring Oscillator With Supply Bus Delay Model

IV. Conclusion

Several fundamentally different type of nonlinear systems that are vulnerable to harboring hardware Trojans have been identified and classified. Trojans in these classes all have the property that they are embedded as undesired solutions of a set of nonlinear differential equations. The Trojan is hidden because the system's desired solution that describes the intended operation of the circuit is more likely. This property makes these Trojans particularly stealthy. Some of the most basic and widely used analog circuits are vulnerable to exposure by these types of Trojans and hence they can be broadly classified as analog hardware Trojans. Defense against these insidious Trojans necessitates a design and verification approach that specifically focuses on developing a better understanding of how nonlinearity introduces multiple modes of operation in a system.

Acknowledgment

This research has been supported, in part, by the Semiconductor Research Corporation and the National Science Foundation.

References

[1] Q. Wang, D. Chen, and Randall. L. Geiger, "Transparent side channel trigger mechanism on analog circuits with PAAST hardware Trojans," in *2018 IEEE International Symposium on Circuits and Systems (ISCAS)*, May 2018, pp. 1–4, doi: 10.1109/ISCAS.2018.8351233.

[2] Y.-T. Wang, D. Chen, and R. L. Geiger, "Practical methods for verifying removal of Trojan stable operating points," in *2013 IEEE International Symposium on Circuits and Systems (ISCAS)*, May 2013, pp. 2658–2661, doi: 10.1109/ISCAS.2013.6572425.

[3] X. Cao, Q. Wang, R. L. Geiger, and D. J. Chen, "A hardware Trojan embedded in the Inverse Widlar reference generator," in *2015 IEEE 58th International Midwest Symposium on Circuits and Systems (MWSCAS)*, Aug. 2015, pp. 1–4, doi: 10.1109/MWSCAS.2015.7282131.

[4] C. Yan and M. Greenstreet, "Oscillator verification with probability one," in *2012 Formal Methods in Computer-Aided Design (FMCAD)*, Oct. 2012, pp. 165–172.

[5] Q. Wang, R. L. Geiger, and D. Chen, "Hardware Trojans embedded in the dynamic operation of analog and mixed-signal circuits," in *2015 National Aerospace and Electronics Conference (NAECON)*, Jun. 2015, pp. 155–158, doi: 10.1109/NAECON.2015.7443059.

[6] C. Cai and D. Chen, "Performance enhancement induced Trojan states in op-amps, their detection and removal," in *2015 IEEE International Symposium on Circuits and Systems (ISCAS)*, May 2015, pp. 3020–3023, doi: 10.1109/ISCAS.2015.7169323.

[7] R. Geiger, "Parasitic pole approximation techniques for active filter design," *IEEE Trans. Circuits Syst.*, vol. 27, no. 9, pp. 793–799, Sep. 1980, doi: 10.1109/TCS.1980.1084896.