
CSIT
CENTRE
FOR SECURE
INFORMATION
TECHNOLOGIES
DEEPSECURITY: APPLYING
DEEP LEARNING TO
HARDWARE SECURITY
QUEEN'S
UNIVERSITY
BELFAST
ESTP1845

DeepSecurity: Applying Deep Learning to Hardware Security

Overall Goal

To investigate the use of Deep Learning for security verification in EDA tools, specifically in relation to *Hardware Trojan detection* and *Side channel analysis* to allow non-security experts to receive feedback on how to improve the security of their designs prior to fabrication.

Rambus Cryptography Research

riscure

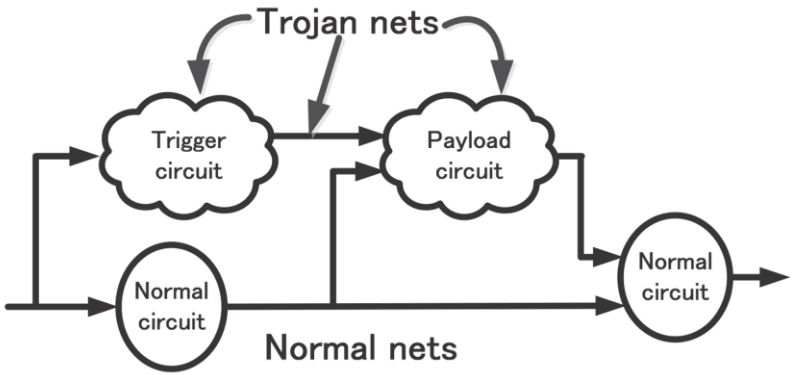
BAE SYSTEMS

CSIT CENTRE FOR SECURE INFORMATION TECHNOLOGIES

Hardware Trojans

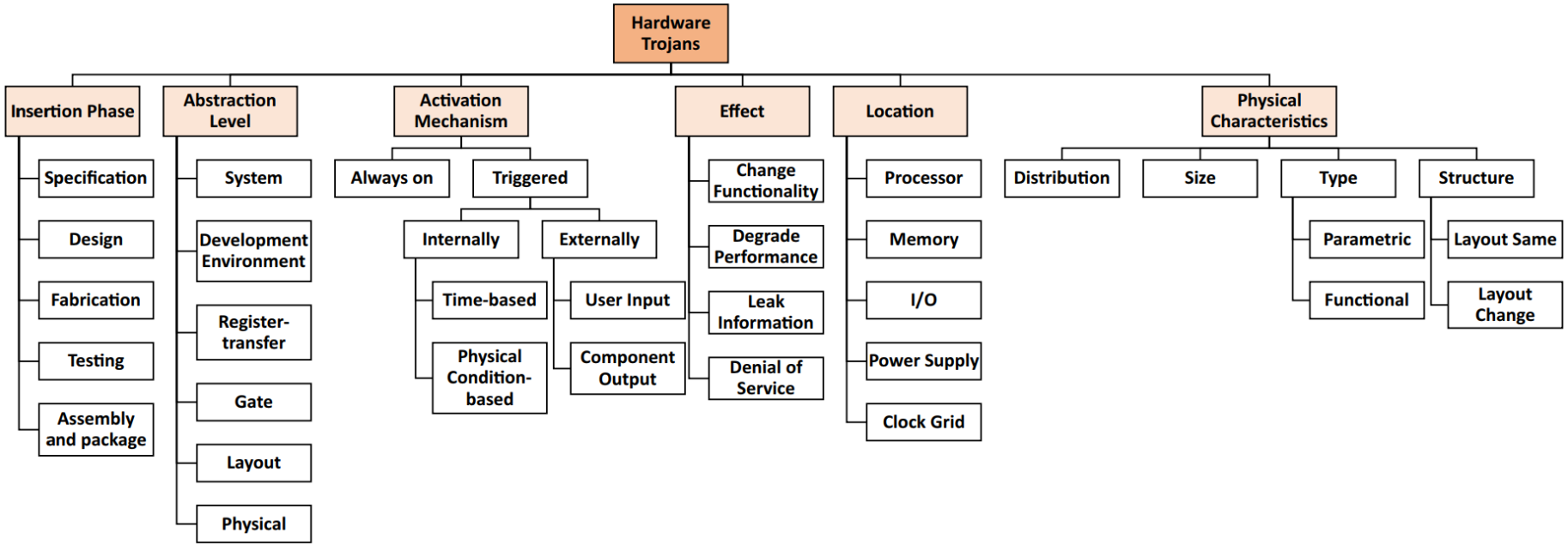
CSIT CENTRE FOR SECURE INFORMATION TECHNOLOGIES

- Hardware Trojans (HTs) are malicious modifications to integrated circuits (ICs)
  - Emerging security concern in the IC industry
- Globalisation of semiconductor supply chains, design and fabrication of ICs now distributed worldwide
  - use of overseas foundries,
  - third party IP,
  - third party test facilities
- Becoming difficult to ensure the integrity and authenticity of devices



HT Classification

CSIT CENTRE FOR SECURE INFORMATION TECHNOLOGIES

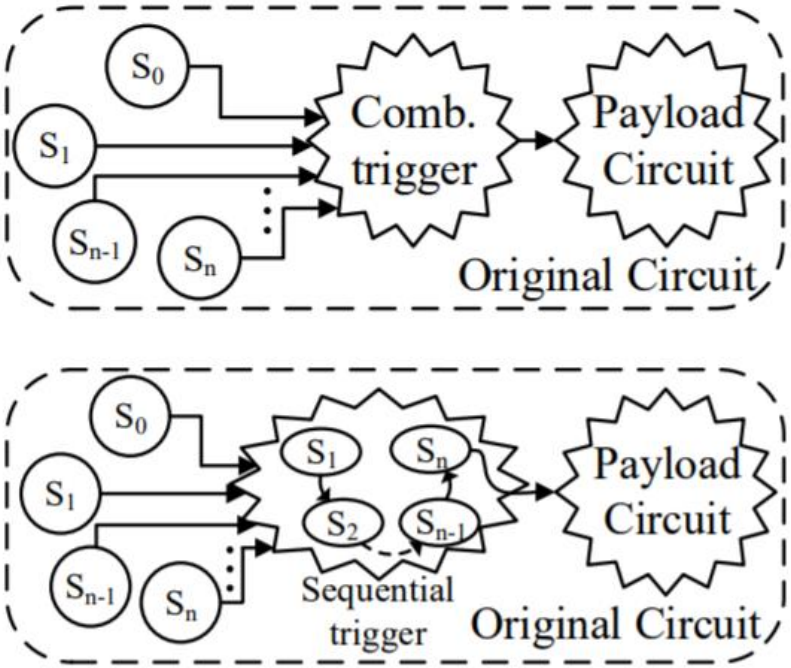


***Source*:** B. Shakya, T. He, H. Salmani, D. Forte, S. Bhunia, and M. Tehranipoor, 'Benchmarking of Hardware Trojans and Maliciously Affected Circuits', *J. Hardw. Syst. Secur.*, pp. 1–18, 2017.

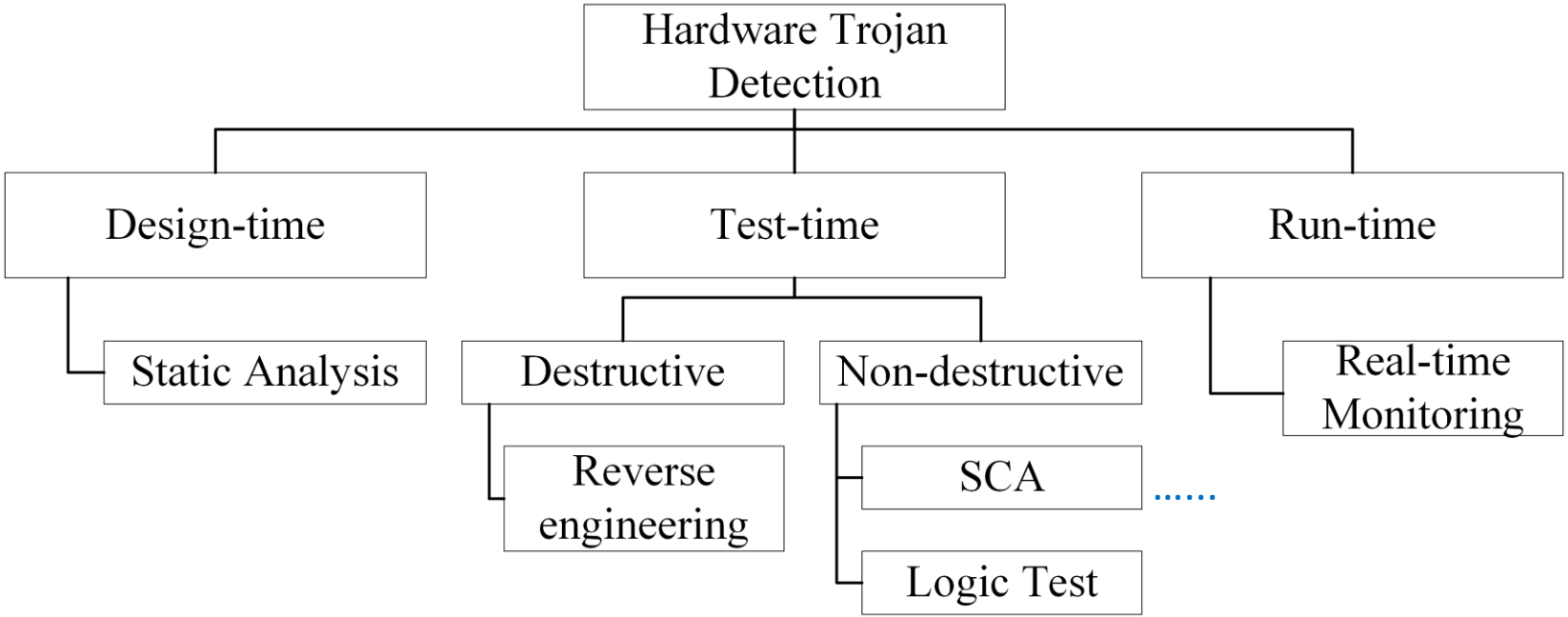
Simple Classification

CSIT CENTRE FOR SECURE INFORMATION TECHNOLOGIES

- Functional Trojans
  - Combinational Trojans
  - Sequential Trojans
- Parametric Trojans
  - Dopant area
  - Doping concentration



Detecting Hardware Trojans – IC Production Stage



Use of Machine Learning to detect HTs

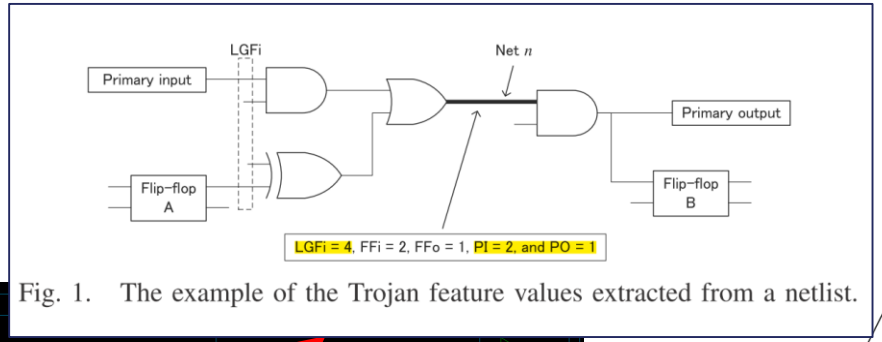
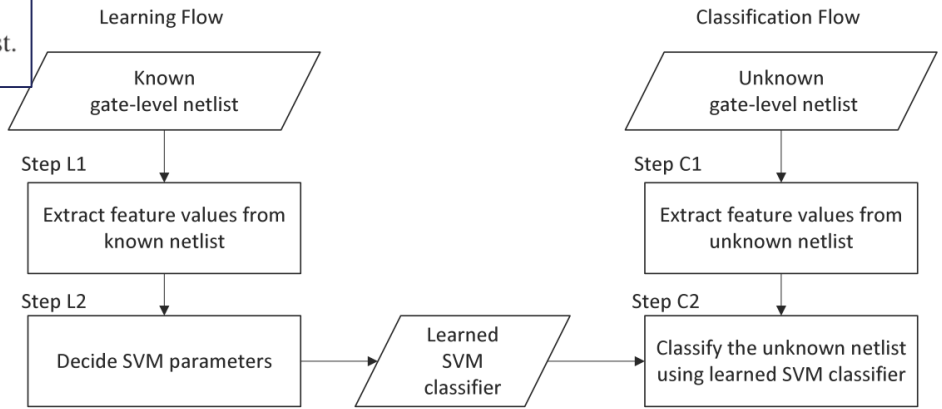


Fig. 1. The example of the Trojan feature values extracted from a netlist.

AVERAGE VALUES OF FFI, FFO, PI, AND PO.

Trojan/Normal	FFi	FFo	PI	PO
Trojan nets	1.11	1.51	4.60	3.65
Normal Nets	0.18	0.16	1.50	1.35

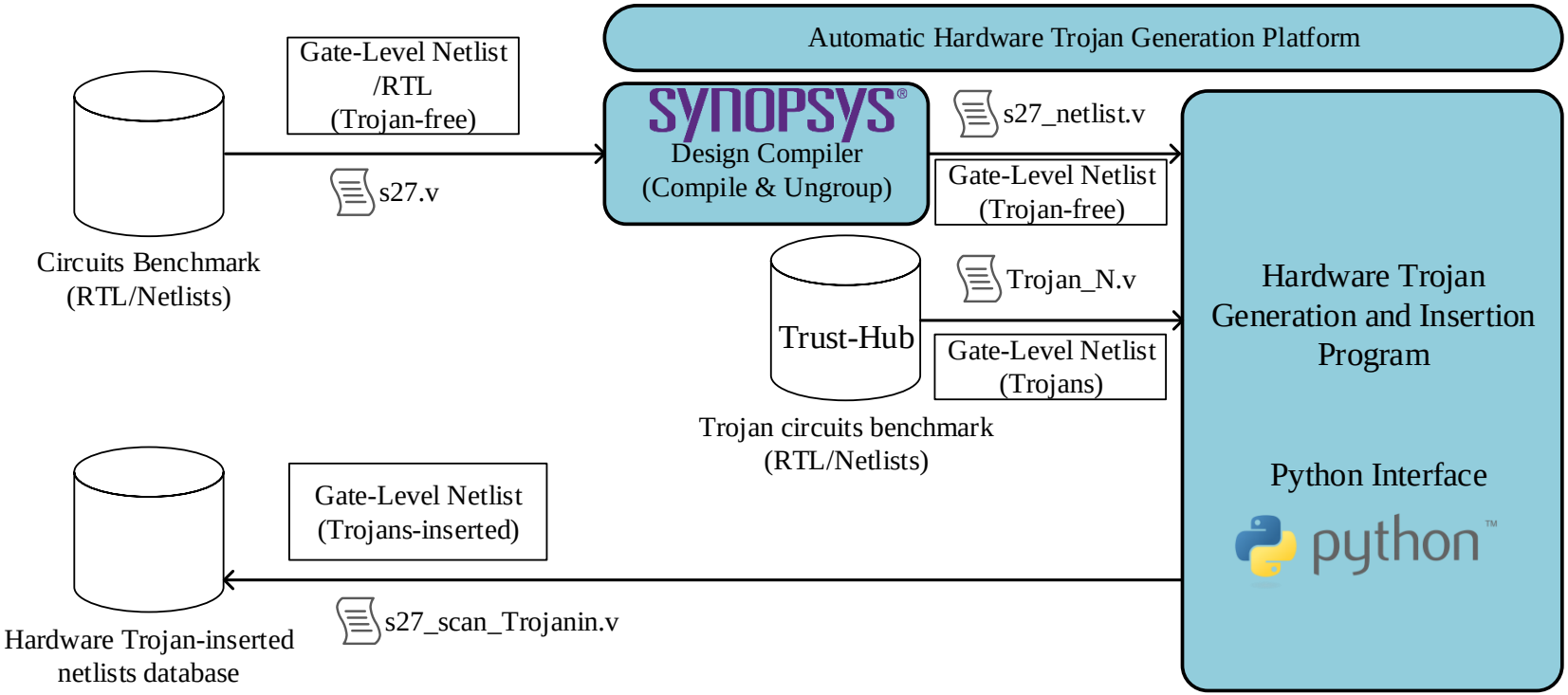
Data	TPR in [1]	TPR in ours	TNR in [1]	TNR in ours
s15850-T100	61%	93%	99%	66%
s35392-T200	27%	100%	99%	59%
s38417-T100	100%	100%	99%	76%
s38584-T200	99%	94%	98%	64%



***Source*:** K. Hasegawa, 'Hardware Trojans classification for gate-level netlists based on machine learning', IEEE 22nd IOLTS, 2016.

Current progress: An Automatic Hardware Trojan Generation Platform-1/2

CSIT CENTRE FOR SECURE INFORMATION TECHNOLOGIES



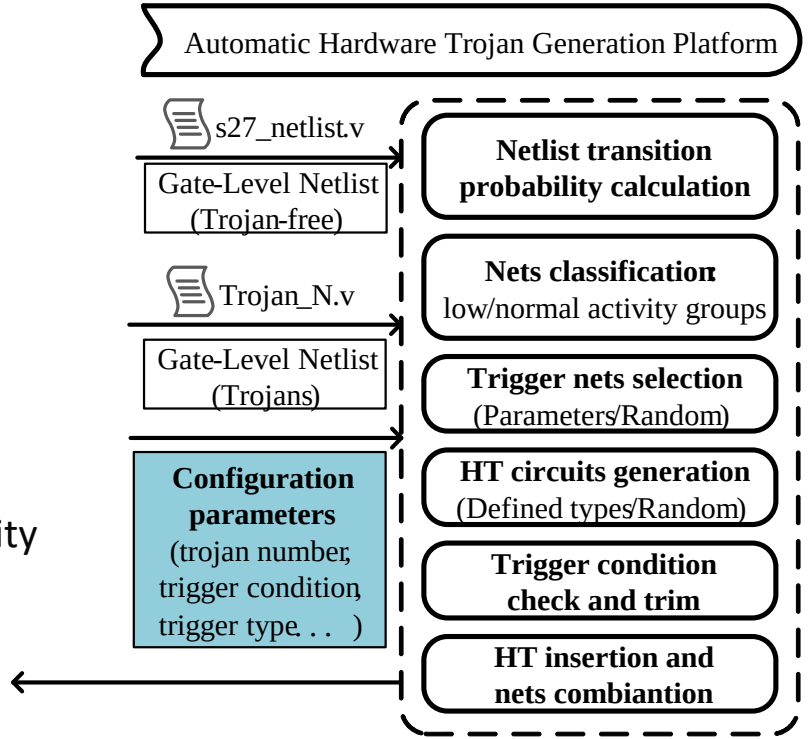
Current progress: An Automatic Hardware Trojan Generation Platform-2/2

CSIT CENTRE FOR SECURE INFORMATION TECHNOLOGIES

Sample of HT Configuration Parameters

Parameters	Value Type
No. of Hardware Trojans	Number
No. of low activity nets	Number
No. of normal activity nets	Number
HT trigger type	Combinational/Sequential
HT payload type	Functional/Leakage
User-designed HT	Enable/Disable
……	……

- Trigger selection according to netlist transition probability
- Highly configurable
- A variety of HT types and benchmark circuit choices
- Pre-research on ML/DL-based HT detection



Next Steps

CSIT CENTRE FOR SECURE INFORMATION TECHNOLOGIES

- Evaluate ML approaches to check the validity of the proposed platform and HT-infected circuit database & make the database publicly available.
- Compare several ML algorithms to determine the best detection approach.
- Step into deep learning, investigate the performance between different NN architectures and build a novel deep learning architecture for HT detection.
- Conduct a comprehensive evaluation of the application of supervised and unsupervised ML and deep learning techniques for HT detection on gate-level netlists.

Thank you