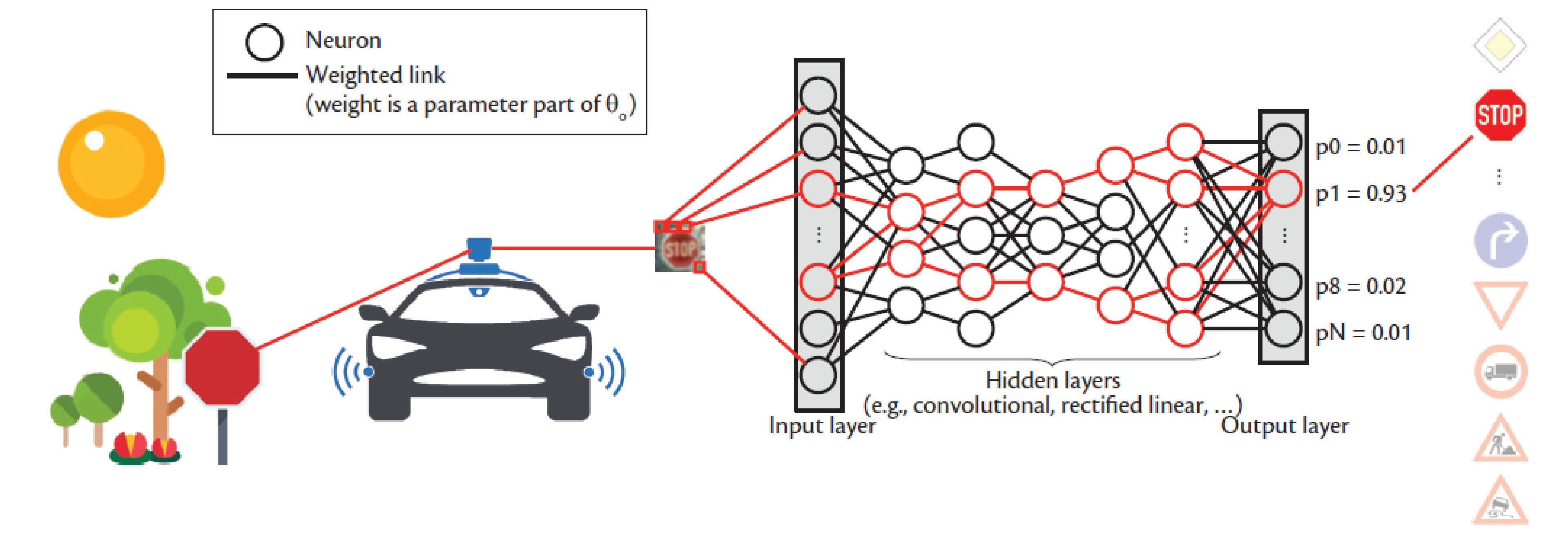
SaTC: Core: Medium: Protecting Confidentiality and Integrity of Deep Neural Networks against Side-channel and Fault Attacks

PIs: Yunsi Fei, Xue Lin, Thomas Wahl, Northeastern University, Boston

Email: yfei@ece.neu.edu; http://nueess.coe.neu.edu, https://nsfchest.org



Medical Image Analysis with Deep Learning

Challenge:

- Serious security implication of DNNs
  - Integrity/availability of DNN execution will affect safety, access control, and trust of DNN-supported systems and services
  - Data is valuable and therefore trained DNN model becomes IP, privacy issue too
  - Existing DNN security does not address side-channel leakage and fault injections
- Diverse models, platforms, optimizations, and applications for DNNs

Proposed Solution: ensure secure DNN execution

- *SpyNet*: recovering DNN structure and parameters on diverse platforms
- *DisruptNet*: manipulate DNN operations via hardware and software fault injections
- *SecureNet*: network obfuscation against side-channel attacks, detection of integrity violation of DNNs, and hardening techniques for fault resistance

Scientific Impact:

- Investigate a new attack surface of DNN inference
- Systematically protect confidentiality and integrity of DNNs
- Deepen understanding of inherent information leakage and fault tolerance of DNN models
- Incorporate formal methods for integrity violation detection and prevention

	Model Information	HW Implementation	SW Implementation
Structure characteristics	# layers	power SPA	
		memory access	μarchitecture (I$, PMC)
	type of layer, activation	power SPA, timing	
	connection/ layers		
Hyperparameters	# neurons in FC	power SPA	μarchitecture (I$, PMC)
	# of kernels in CONV	power SPA, memory access	I$, D$, PMC, constraints
	size of kernel in CONV/POOL	memory access	I$, D$, PMC, constraints
Parameters	weights in FC	power DPA, bus snoop	FP
	kernel values in CONV	power DPA, bus snoop	timing μarchitecture, FP

		HW implementation		SW implementation	
		Resource	Fault Type	Resource/Stage	Fault Type
Computation		datapath PE	output: stuck-at, random	instruction execution	skip,control/data flow
		control logic	control flow		
Data	reuse	buffer	set/reset, random	DRAM	set/reset, random, flip (rowhammer)
	temporary	registers	set/reset, random (DVFS)	registers	set/reset, random (DVFS)

Broader Impact:

- Facilitate wide adoption of DNN in security-critical applications
- Advance the state-of-the-art DNN implementations (edge and cloud), heterogeneous systems, hardware security, formal methods and verification
- Interdisciplinary research

Education and Outreach:

- Integrate research with education – hardware security, DL implementation, engineering reliable software
- Engage underrepresented students, undergraduates, high-school students in research
- Technology transfer with company partners through a new NSF IUCRC center

Assessment:

- 9 undergraduates from partner community colleges and HBCUs
- 6 high school seniors from URM groups through NU Young Scholars Program
- Professional from Programs and Operations in the Center of STEM Education to assist outreach and assessment

The 4th NSF Secure and Trustworthy Cyberspace Principal Investigator Meeting (2019 SaTC PI Meeting)
October 28-29, 2019 | Alexandria, Virginia