Gotcha! I Know What You Are Doing on the FPGA Cloud: Fingerprinting Co-Located Cloud FPGA Accelerators via Measuring Communication Links

Chongzhou Fang, Ning Miao, Han Wang, Jiacheng Zhou, Tyler Sheaves, John M. Emmert, Avesta Sasan, Houman Homayoun

Nov. 28, 2023

Introduction

Motivation

- FPGAs have been widely deployed in public clouds for computation acceleration
  - AWS
  - Alibaba Cloud
  - …
- Application Scenario
  - Simulation [1]
  - AI inference [2]
  - …

[1] Karandikar, Sagar, et al. ”FireSim: FPGA-accelerated cycle-exact scale-out system simulation in the public cloud.” ISCA’18.
[2] Wang, Chao, et al. "DLAU: A scalable deep learning accelerator unit on FPGA." IEEE TCAD, 36.3(2016):513-517.

Motivation

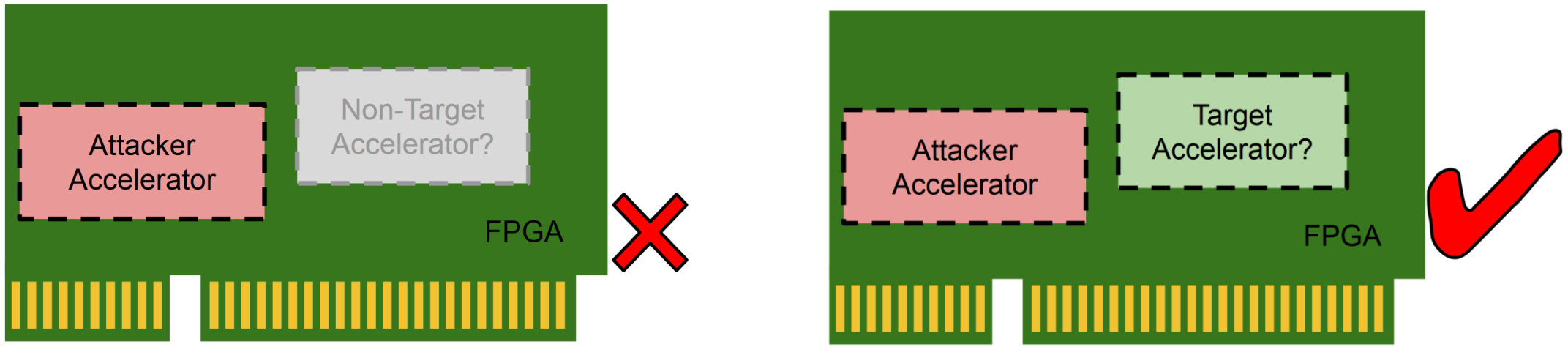
- Security Concerns
  - Side-Channel Attacks
    - Long-wire Side-Channel Attack [3]
    - Power Side-Channel Attack [4]
    - …
- Shared FPGA resources

- Prerequisite of Attacks
  - Co-location and co-location verification of attacker and victim

[3] Giechaskiel, Ilias et.al. "Leaky wires: Information leakage and covert communication between FPGA long wires." AsiaCCS'18.
[4] Zhao, Mark, et.al. "FPGA-based remote power side-channel attacks." S&P'18.

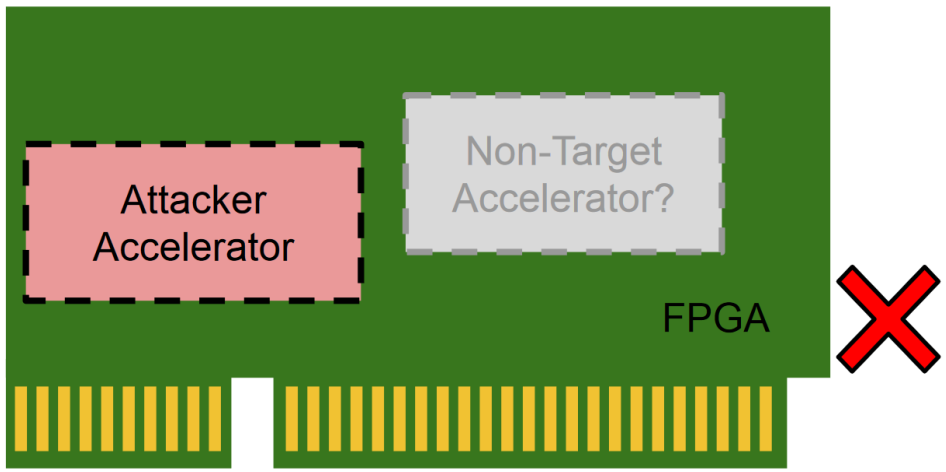
Contributions

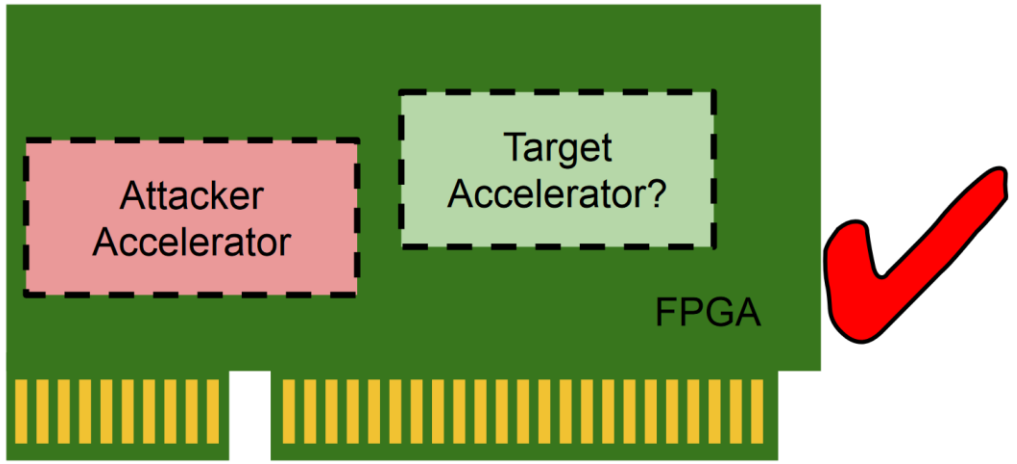
- An FPGA Accelerator Fingerprinting Method



Threat Model

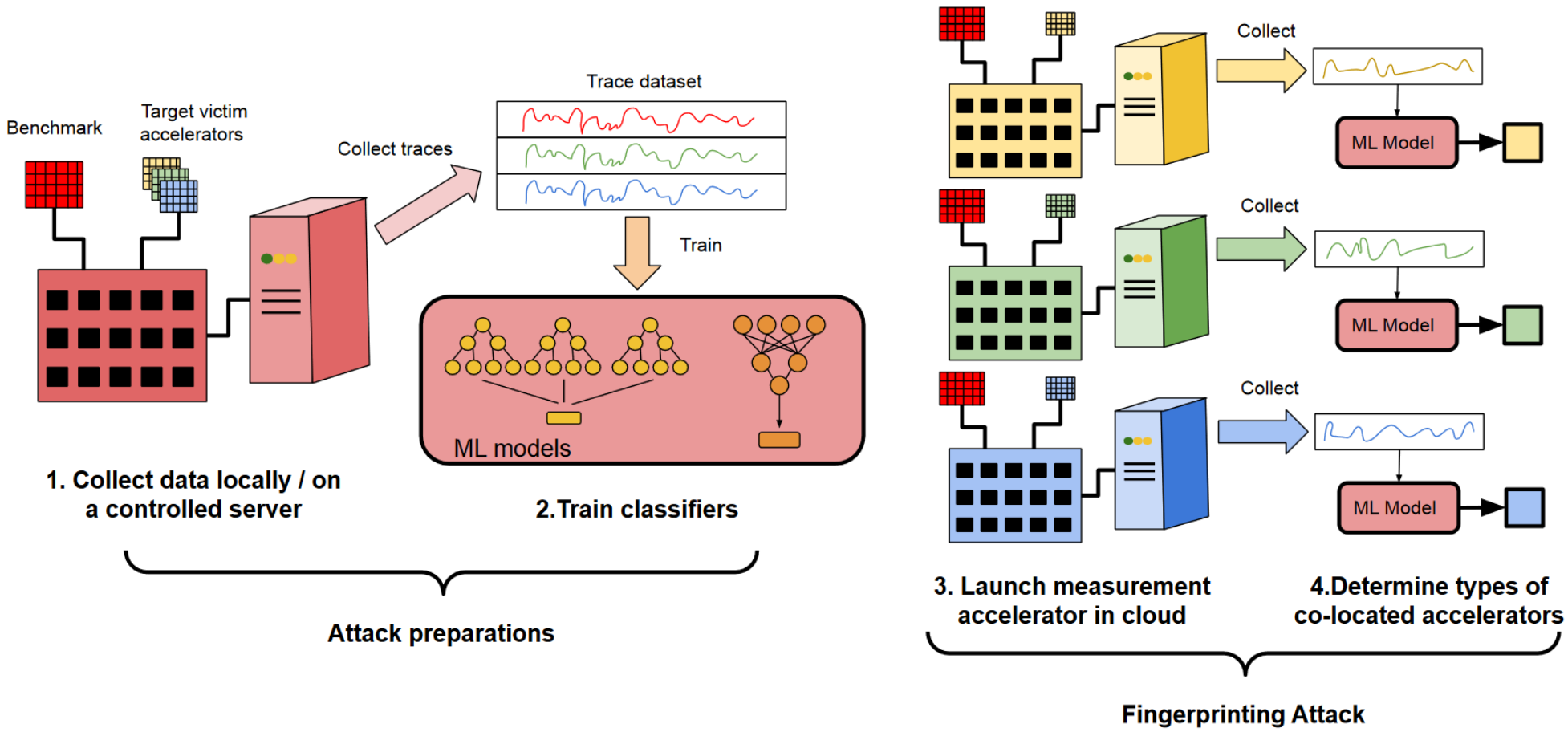
- Focus: FPGA instances on the same board
- Assumption: Both victims and attackers are cloud FPGA users without any privilege





Method and Implementation

Method


Benchmark
Target victim accelerators
Collect traces
Trace dataset
Train
ML models
1. Collect data locally / on a controlled server
2.Train classifiers
Attack preparations
Collect
ML Model
Collect
ML Model
Collect
ML Model
3. Launch measurement accelerator in cloud
4.Determine types of co-located accelerators
Fingerprinting Attack

Design

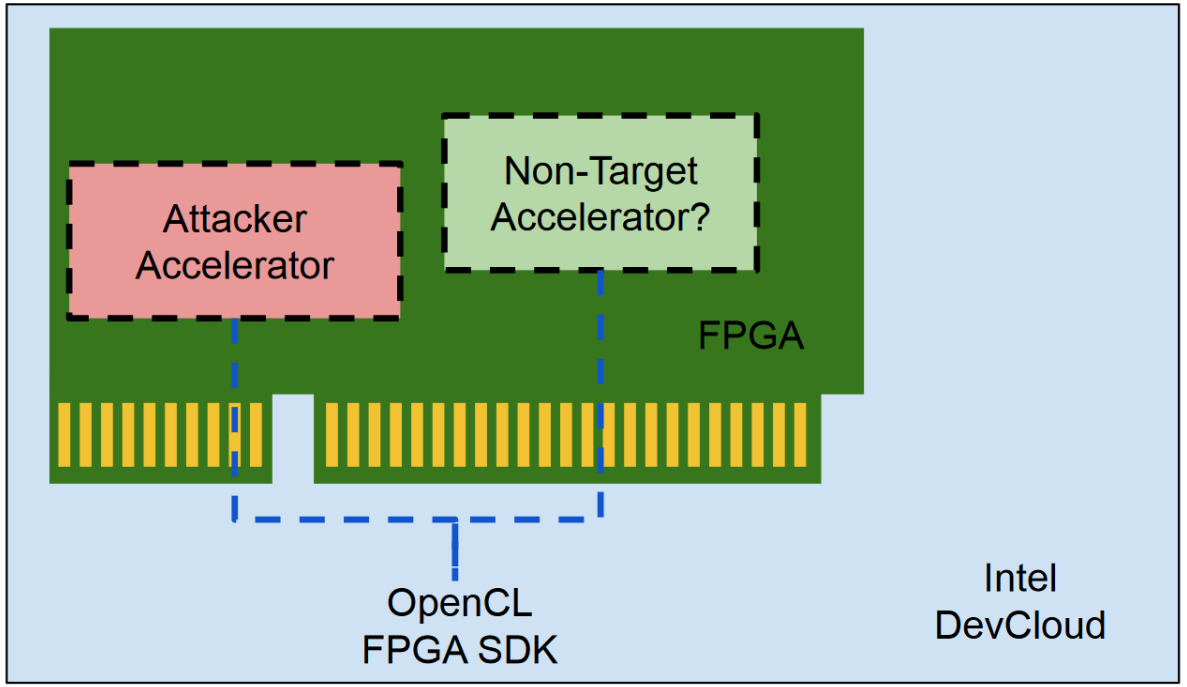
- PoC Design

Host

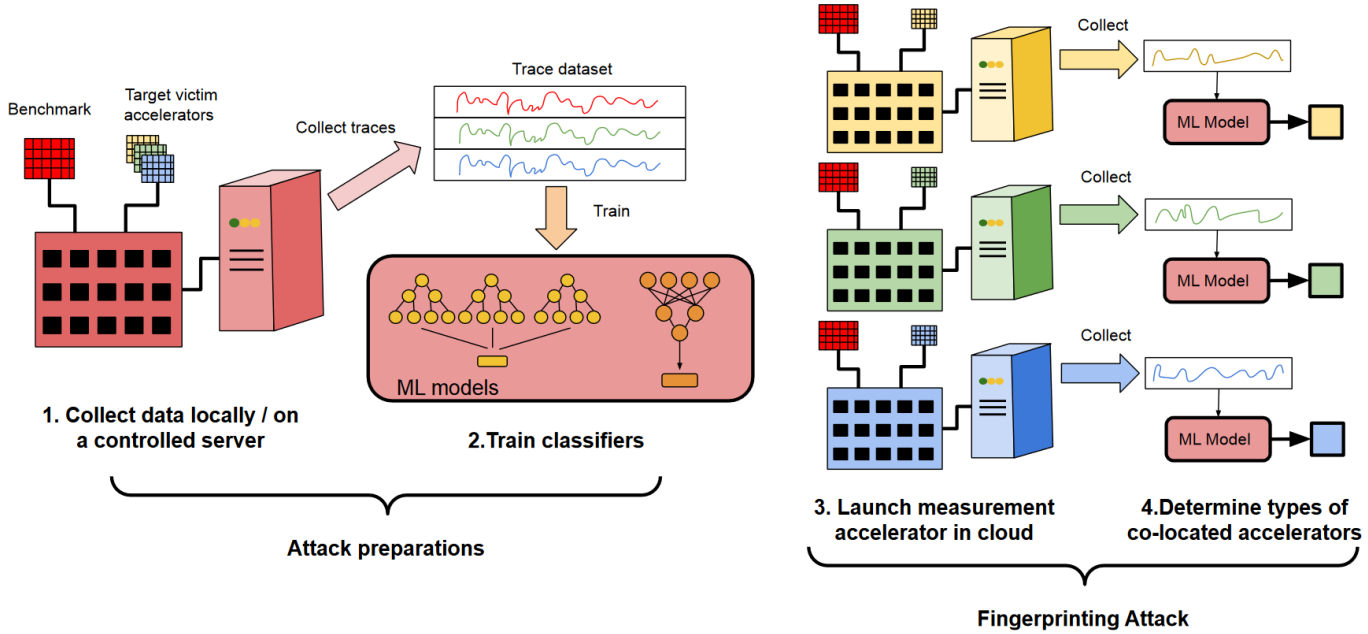
```
Allocate buffer[BUFFER_SIZE][1..BUFFER_NUM];
trace = [];
for(i = 1..BUFFER_NUM) {
  t_i = 0;
  for(j = 1..REPEAT_NUM) {
      call accelerator and operate on buffer[i];
      t_i += time of kernel execution;
  }
  t_i /= REPEAT_NUM;
  trace.append(1 / t_i);
}
return trace;
```

Accelerator

```
__kernel void mem_kernel(__global int4 *dst) {
    int id = get_global_id(0);
    for(long i = 0; i < ACCESS_NUM; i ++) {
        dst[id] = (int)dst ^ dst[id];
    }
}
```



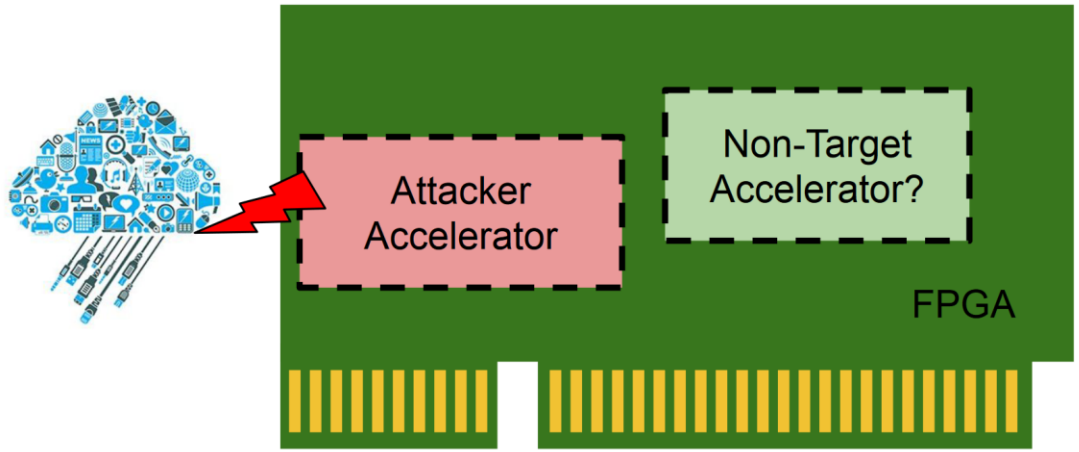
Method



- Collect data locally / on a controlled server in cloud
- Train classifiers
- Launch attack
- Online classification

Evaluation

Research Questions (RQs)



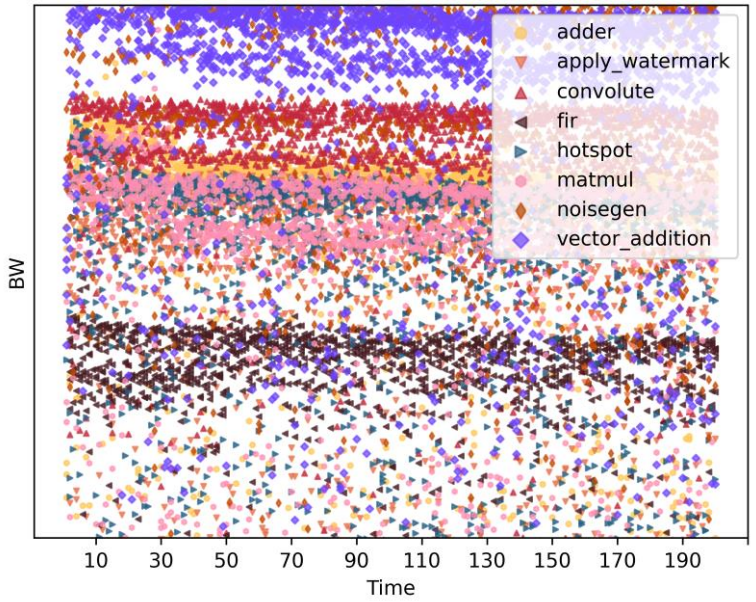
- *RQ1*: Does our attack circuit capture the communication patterns, and what is the accuracy of fingerprinting?
- *RQ2*: How do the parameter settings of our attack impact the attack results?

Experiment Settings

- Server
  - `s005-n007` on Intel DevCloud
  - OpenCL FPGA SDK
- Targeted FPGA Accelerators

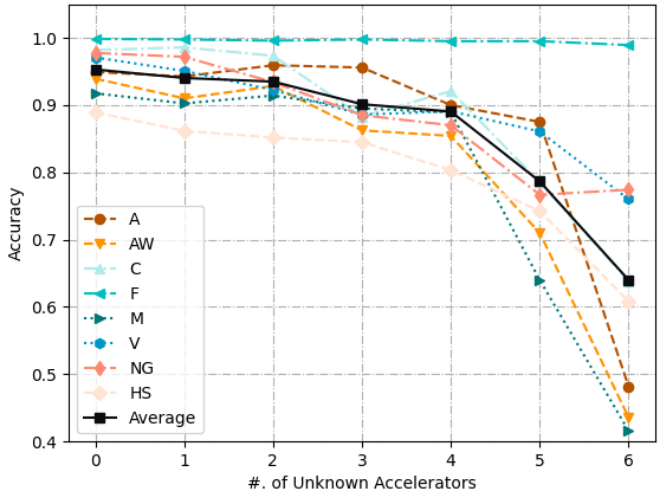
Name	Code	Function
`adder`	`A`	Adder implemented using FPGA. It reads inputs from input buffer, computes results and writes back to a output buffer.
`apply_watermark`	`AW`	Image processing circuit. It reads an image from input buffer, adds a watermark andhiheh hack to a output buffer.
`fir`	`F`	Signal processing circuit. It reads input and coefficient data from input buffer and performs finite impulse response (FIR) filtering, then writes output back to output buffer.
`matmul`	`M`	Matrix multiplication circuit. It reads two matrice A and B from input buffer, calculates AB and writes back to output buffer.
`convolute`	`C`	Convolution accelerator. It reads an image and filter weights from input buffer, performs convolution and writes the results back to output buffer.
`vector_addition`	`V`	This accelerator reads two arrays from input buffer, performs parallel vector addition on the two buffers and writes the results back to output buffer.
`noisegen`	`NG`	An accelerator that generates random traffic between host and FPGA.
`hotspot`	`HS`	An accelerator employed from Rodinia benchmark **?** that performs thermal simulation by iteratively solving differential equations.

Answers to RQ1: Trace Visualization

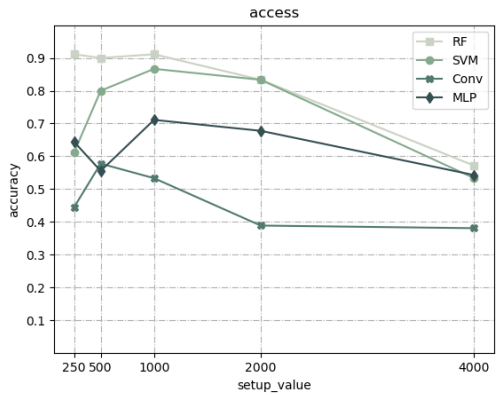


Answers to RQ1: Classification Accuracy

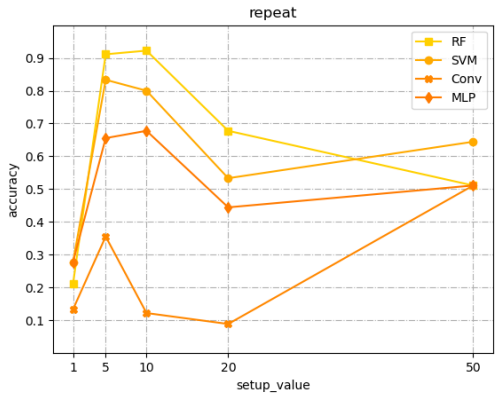
Model	Test Acc.
Random Forest	88%
SVM	69%



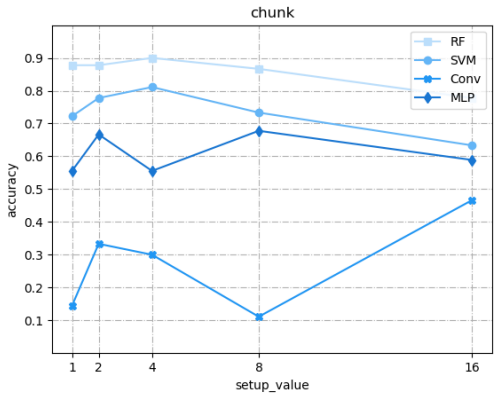
Answers to RQ2: Accuracy Impacts



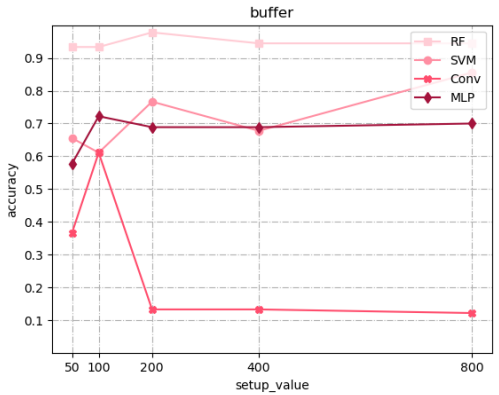
Varying ACCESS_NUM.



Varying REPEAT_NUM.



Varying BUFFER_NUM.



Varying BUFFER_SIZE.

Discussion & Conclusion

Mitigation

- Enhancing FPGA-CPU Communication Interfaces
  - Software
  - Hardware

Conclusion

- An FPGA fingerprinting attack
- Existence of communication side-channel

Q&A