SmartSystems Lab

UF Herbert Wertheim College of Engineering
Department of Electrical & Computer Engineering
UNIVERSITY of FLORIDA

DOMAIN ISOLATION IN FPGA-ACCELERATED CLOUD AND DATA CENTER APPLICATIONS

Joel Mandebi, Sujan Kumar Saha, and Christophe Bobda

Date: June 24, 2021

GLSVLSI 2021

Joel Mandebi

Sujan Saha

Dr. Christophe Bobda

Smart System Lab

- https://smartsystems.ece.ufl.edu/
- https://scholar.google.com/citations?hl=en&user=9uPi2JQAAAAJ
- https://scholar.google.com/citations?user=cgtLTMEAAAAJ&hl=en

Motivation

FPGAs are making their way to Cloud and data centers

We focus on this use case

- **Amazon Web Service (AWS): provides FPGA-attached VMs** with FPGA CAD tools.

- **PLUNIFY: runs design-space exploration** with machine learning algorithms in the cloud to achieve timing closure and reduce design iterations.

VM : Virtual Machine

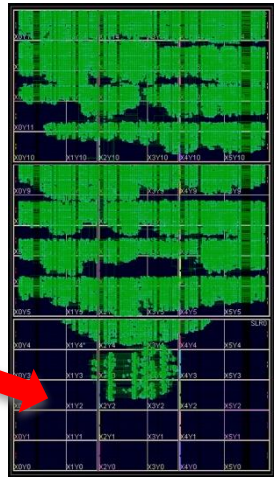
Motivation

Resource waste (FPGA space-sharing is to explore !)

More than 99% of the chip remains unutilized!

An AES128 accelerator only utilizes 0.39% of the chip area on a VU9P

More than 60% of the chip remains unutilized!



A split-CNN architecture only uses 39% of the chip area on VU9P

- A Xilinx **VU9P** has about 2.5 millions logic elements, 6800 DSP slices, 75 MB of BRAM.

↓ 2019

- A Xilinx **VU19P** features 9 millions logic cells

↓

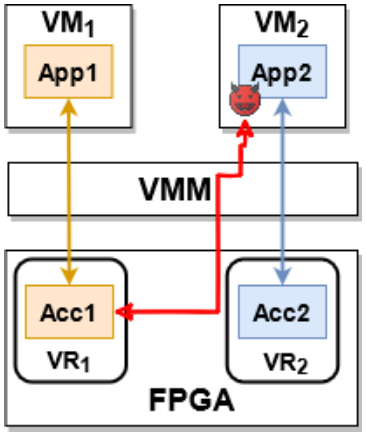
?

Bhowmik, P.,Pantho, M. J. H., Mandebi, J., & Bobda, C. (2021). ESCA: Event-Based Split-CNN Architecture with Data-Level Parallelism on UltraScale+ FPGA. (FCCM). IEEE.

Threat Model

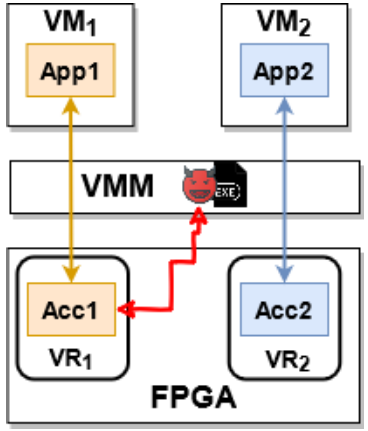
This work address one aspect of security issues of FPGA multi-tenancy: software attempts to breach hardware domain

Scenario1: a VM application accesses the address space of an FPGA accelerator of another VM



Risk of information leakage, task hiding, denial-of-service, etc

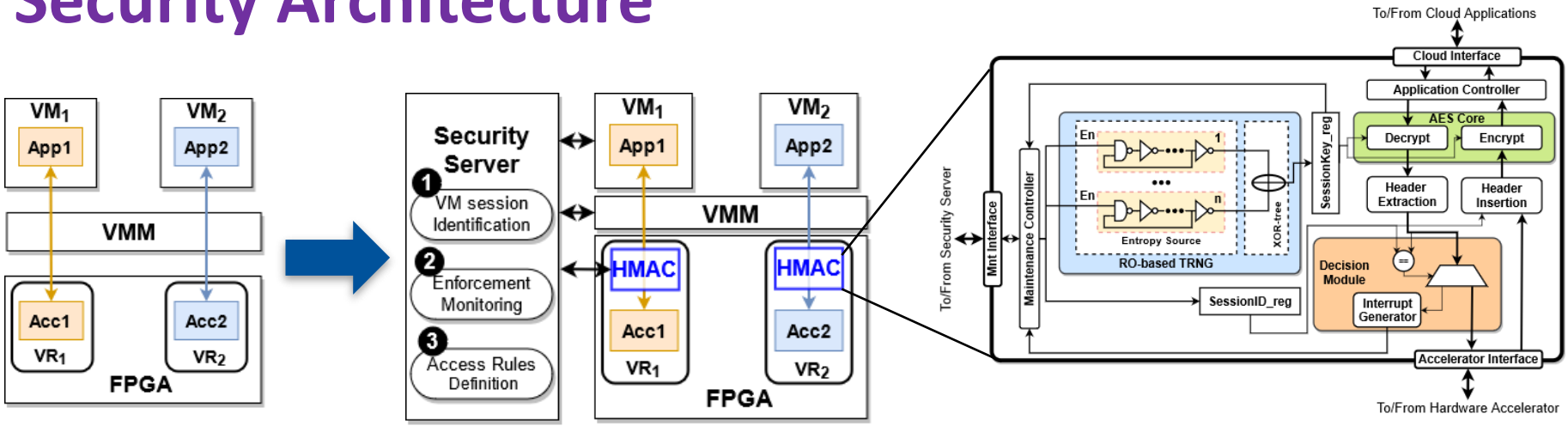
Scenario 2: a software in the VMM host attempts accessing data of FPGA accelerators



VMM : Virtual Machine Monitor

Note: We do not consider attacks initiated from hardware!

Security Architecture



- ✓ The **Security Server** generate random sessionID identifiers that are shared between VM and FPGA accelerator
- ✓ The **Hardware Mandatory Access Controller (HMAC)** generates random session keys using a RO-based true random number generator

Communication between software and hardware accelerators are encrypted, and checked against the correct sessionID

Experimental Evaluation

FPGA Resource Overhead

- **Platform:** Dell EMC R74151 cloud server, Intel Stratix V FPGA, Quartus Prime 18.1.0
- **Experiments:** FPGA resource overhead, Configuration and communication overhead, quality of randomness, case study

- Without considering the host accelerator, a VR uses ~1% of the FPGA area

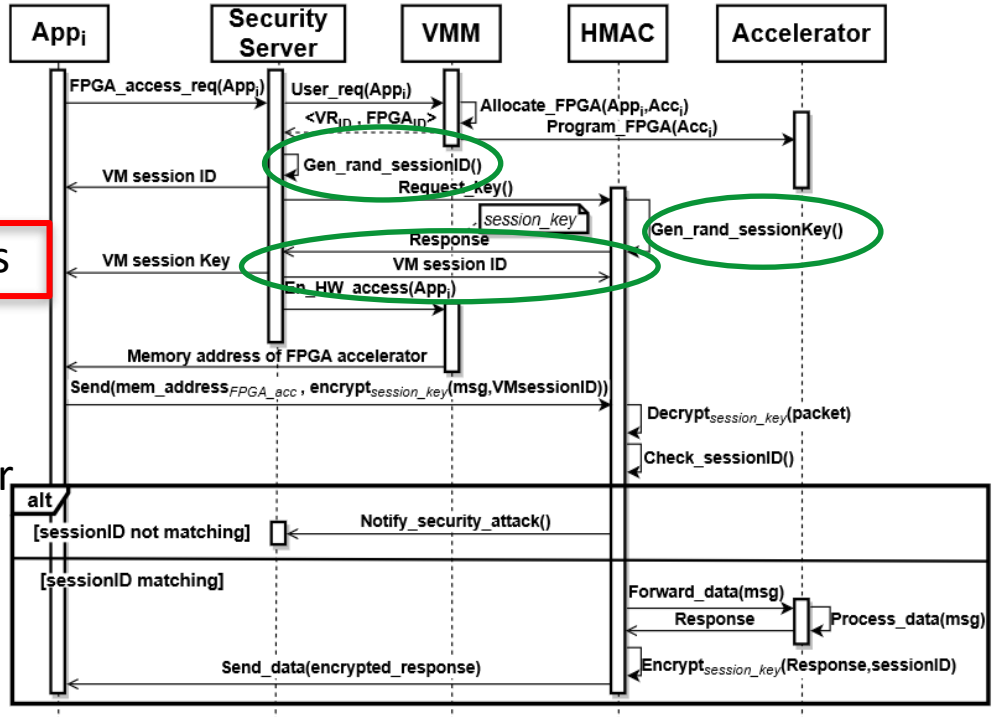
	ALMs	ALUTs	Registers	M20Ks
Total	185000	185000	740000	2100
Used	2830.1	3145	2152	4
Utilization	1.53%	1.7%	0.29%	0.19%
Other Controls	266.4	44	667	0
HMAC	2563.7	3101	1485	4
— TRNG	261	300	275	0
— Decrypt	1213.2	1496	632	4
— Encrypt	1089.5	1305	578	0

Table1. FPGA resource overhead of the VR

Experimental Evaluation

Configuration and Communication Overhead

- Generating a sessionID takes ~10ms
- Generating a session key takes 1.84ns
- Roundtrips between Security Server and HMAC takes ~34µs
- Configuration in the order of milliseconds
- Incoming packets are processed in 14 clock cycles (12 cycles to decrypt, 1 cycle to extract the header, 1 cycle to accept or reject the packet)
- Outgoing packets are processed in 13 clock cycles (1 cycle to insert the header, 12 cycles to encrypt)



Experimental Evaluation

Quality of Randomness

- Implemented 5-stage through 17-stage RO-based TRNGs with 30 ROs, we note. The average hamming distance between the 128-bit session keys is 63.45: in average there are 2^{63} possible numbers between two consecutive keys.

- The average hamming distance between the 16-bit sessionIDs is 8.64: in average there are 2^{8} possible numbers between two consecutive sessionIDs

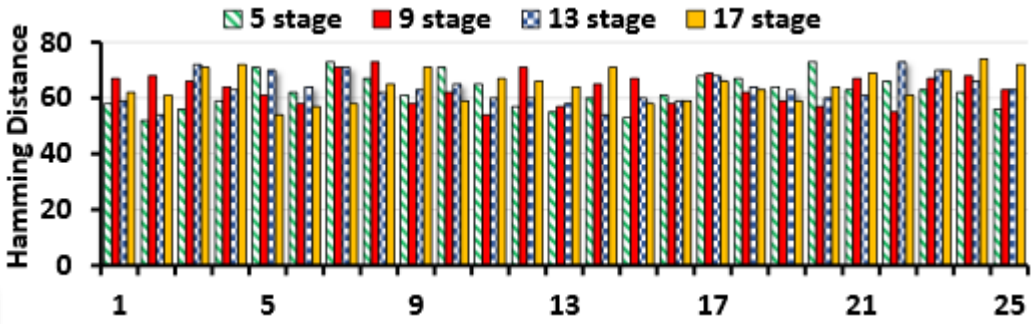


Fig4. Hamming distance between random 128-bit keys

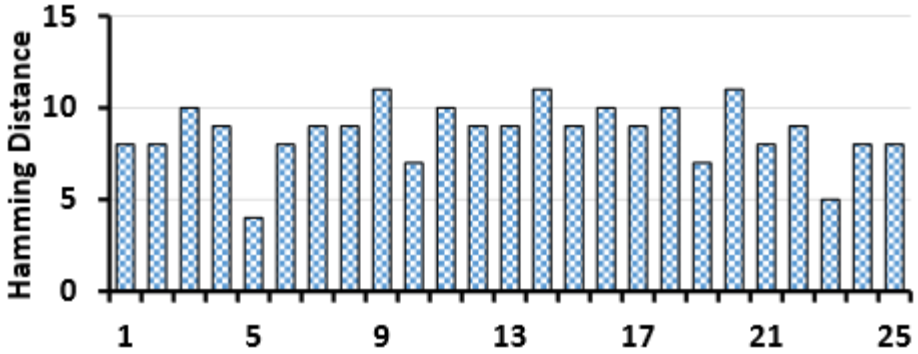


Fig5. Hamming distance between random 16-bit sessionIDs

Experimental Evaluation

Case Study

- A VR implements a JPEG encoder and is assigned to VM1
- VM2 and a malicious application in the VMM try to use base address + offset information of the VR to read/write to the JPEG Encoder
- The HMAC rejects malicious access requests

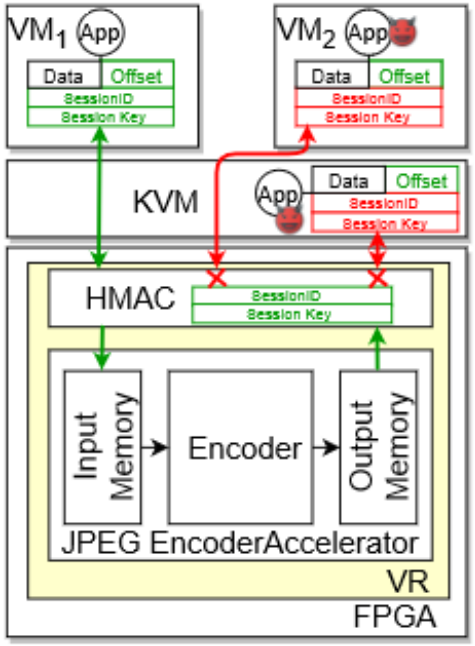


Fig6. Illustration of the domain isolation

Summary

This work proposed an architecture and communication protocol for domain isolation in FPGA-accelerated cloud and data center applications:

- **Secured communication protocol:** it rely on a trusted security server and random session keys generated on FPGA to establish secured and encrypted communication between software running in the VM, and co-hosted hardware accelerators on FPGA
- **Security Architecture:** FPGA accelerator an enabled with the ability to verify that incoming traffic are actually generated by a trusted VM.
- **Overhead:** the proposed isolation mechanism incur minimal overhead in FPGA resource and communication/configuration latency.

THANK YOU

Questions ?

References

- Amazon, Amazon ec2 f1 instances, retrieved February 25, 2021 from https://aws.amazon.com/ec2/instance-types/f1/(2016)
- Alibaba, Cloud server ecs, retrieved February 25, 2021 fromhttps://cn.aliyun.com/price/product#/ecs/detail(2019)
- Amazon, Amazon ec2 pricing, retrieved February 25, 2021 fromhttps://aws.amazon.com/ec2/pricing/on-demand/(2006).
- Huawei, Fpga accelerated cloud server, retrieved February 25, 2021fromhttps://www.huaweicloud.com/en-us/product/fcs.html(2017)
- Huawei, Elastic cloud server, retrieved February 25, 2021 fromhttps://www.huaweicloud.com/en-us/product/ecs.html(2017)
- Nimbix, Nimbix cloud price calculator, retrieved February 25, 2021 fromhttps://www.nimbix.net/cloud-price-calculator#/(2014)

Security Architecture

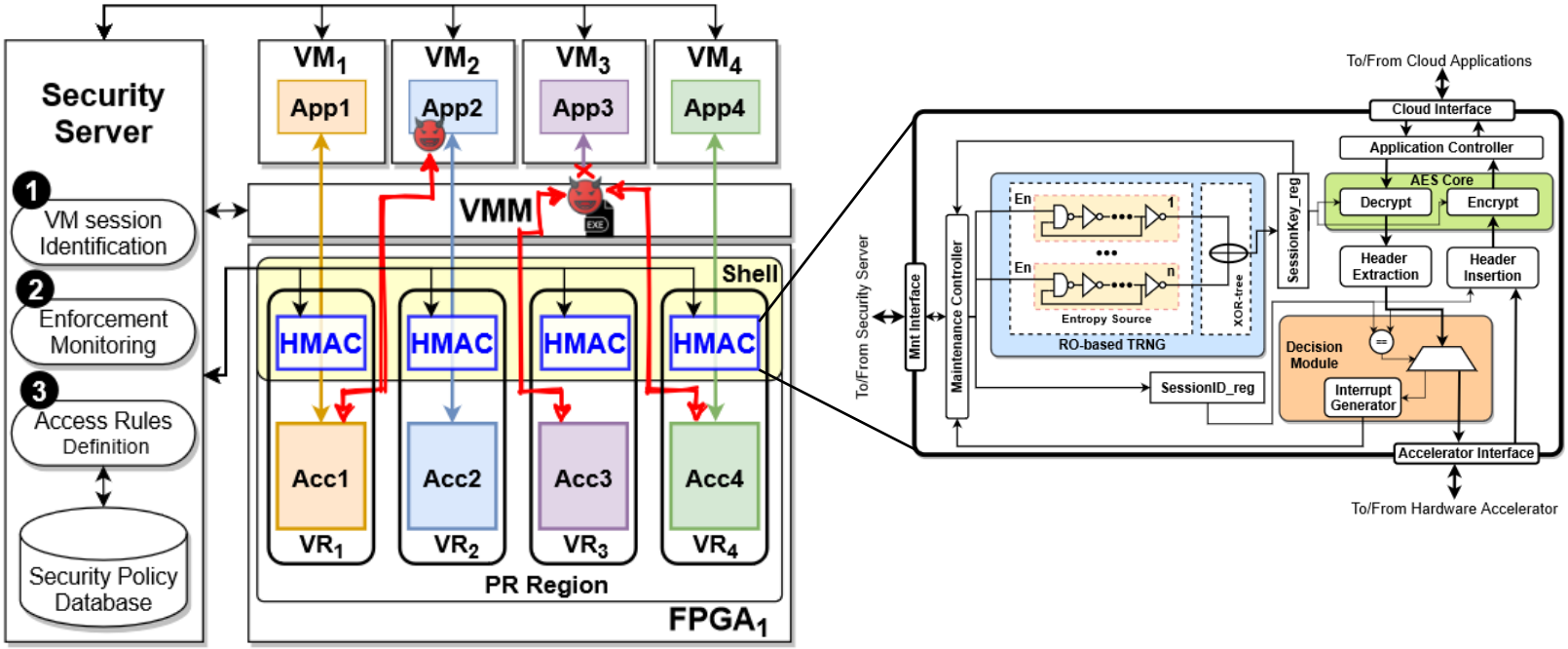


Fig2. Architecture Overview

Secured Communication Protocol

blblbla

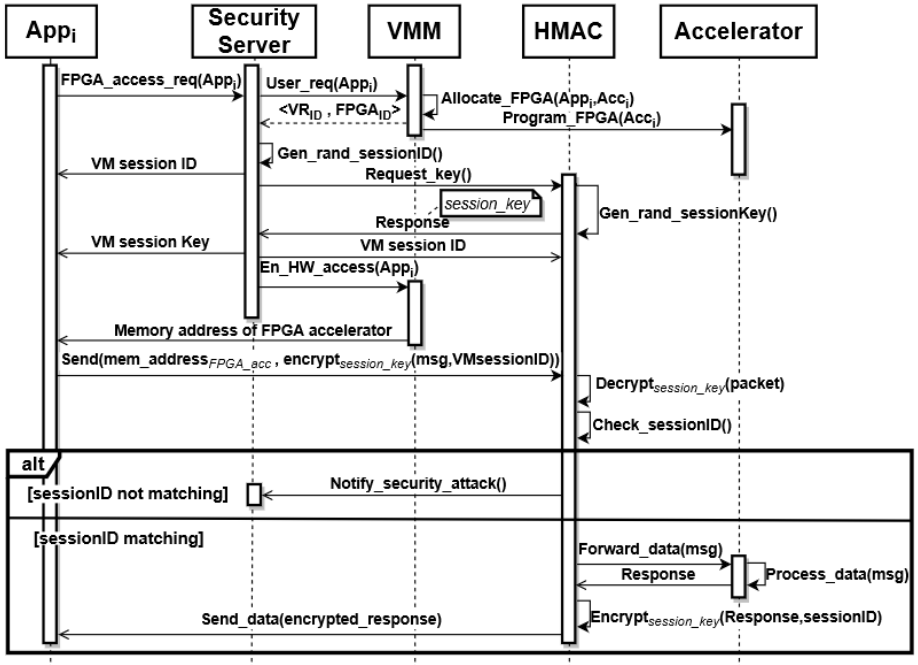


Fig3. Secured communication protocol