Open. Together.

OCP
REGIONAL
SUMMIT

Offloading TLS Onto Crypto SmartNIC
A Technical Introduction

Nic Viljoen, Associate Director of Engineering, Netronome

Introduction: Agenda

- Introduction
- Background
- Firmware
- Driver & Kernel
- Initial Tests
- Performance Analysis
- Summary
- Next Steps

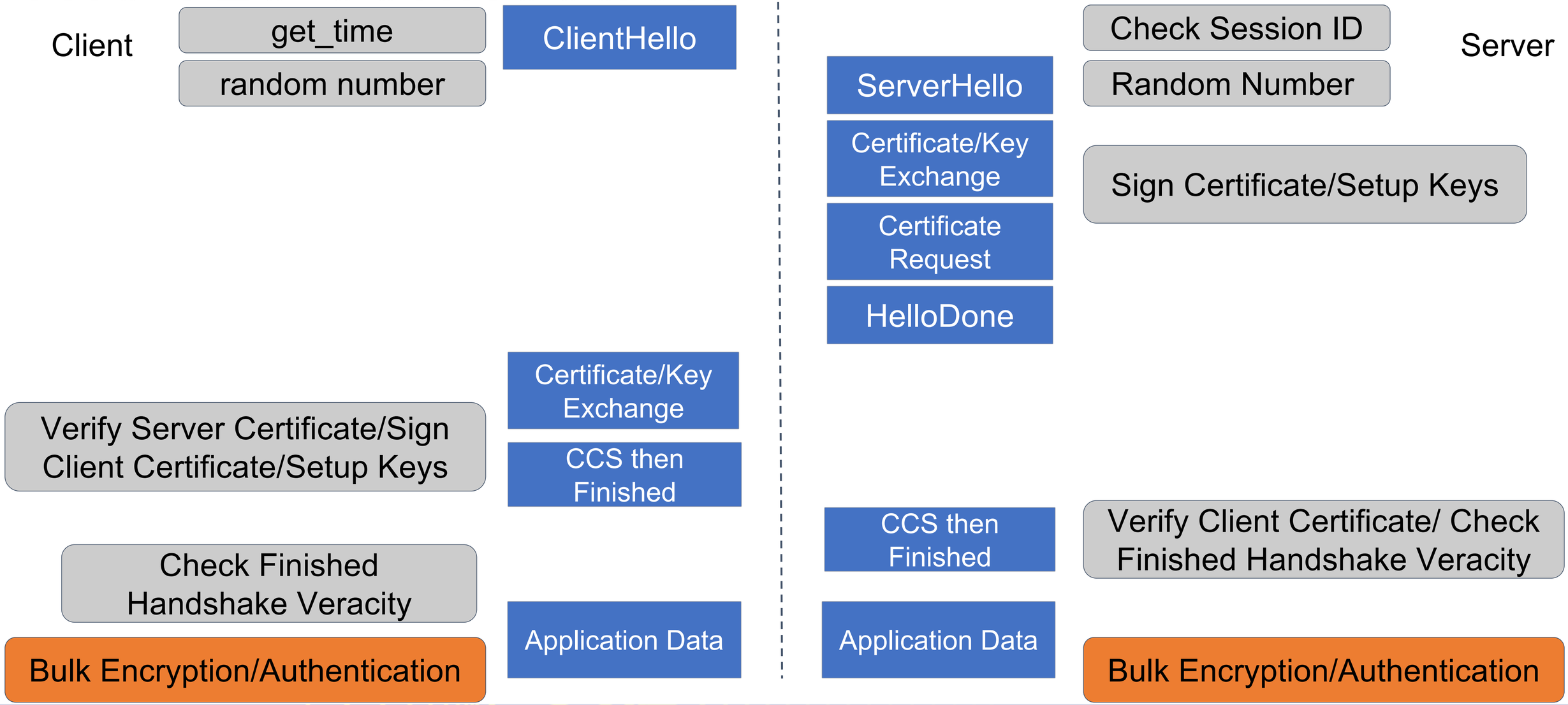
SERVER

Design Files

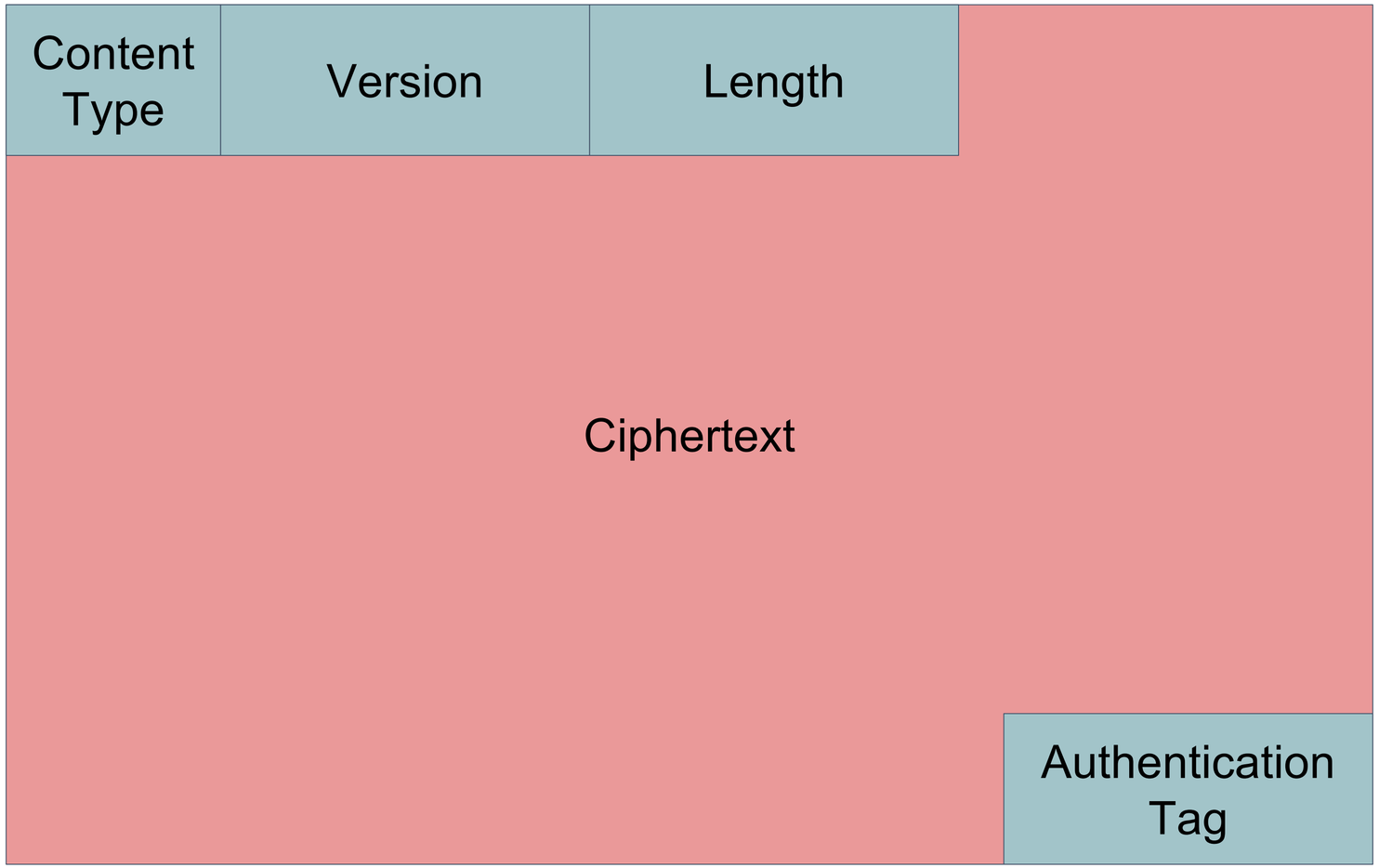
Product Recognition

Introduction to TLS

The TLS Connection



The TLS Record



Content Type (8 bits)

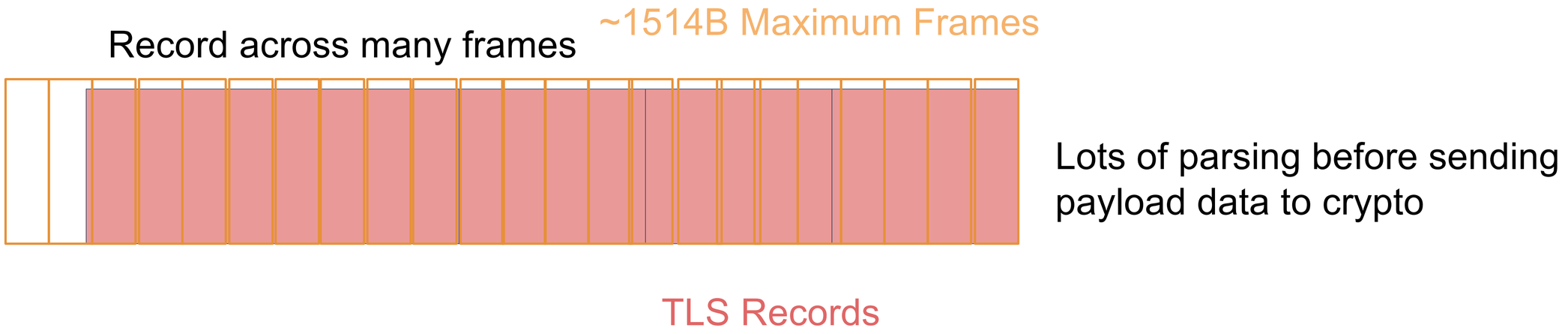
1. Change cipher specification (0x14)
2. Alert (0x15)
3. Handshake (0x16)
4. Application (0x17)

Version (16 bits)

Length (16 bits)

TLS Record can be up to 16KB

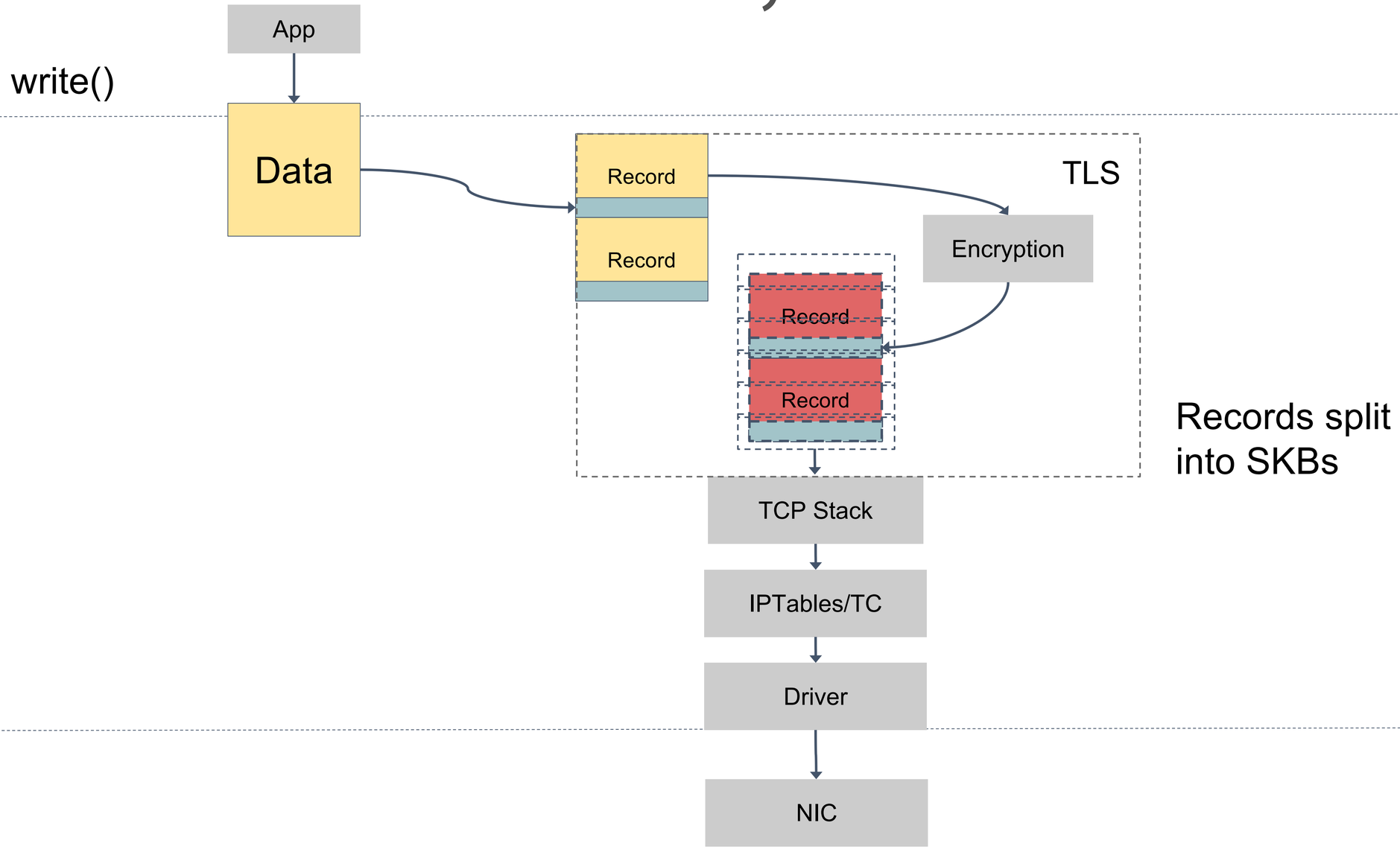
TLS (is always) over TCP



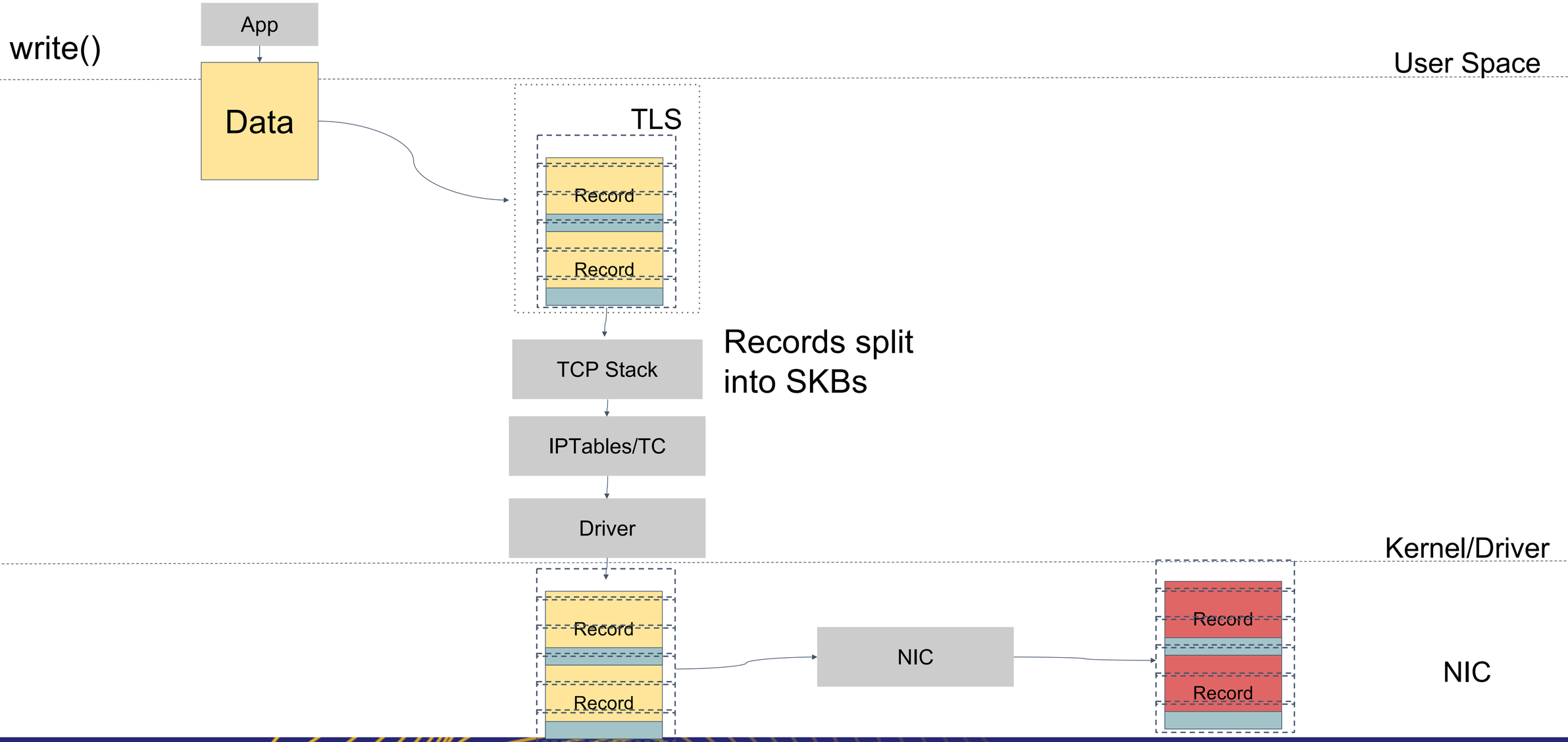
- TLS is designed to be handled *in order*
- However, the exact implementation details are cipher dependent
  - AES-GCM makes this slightly more flexible
- KTLS is upstream in kernel TLS processing handling TLS as a TCP ULP
  - ULP-Upper Layer Protocol
  - KTLS can be exploited for offload
- Everything in this presentation is done with upstream Linux

Kernel TLS & Offload Architecture

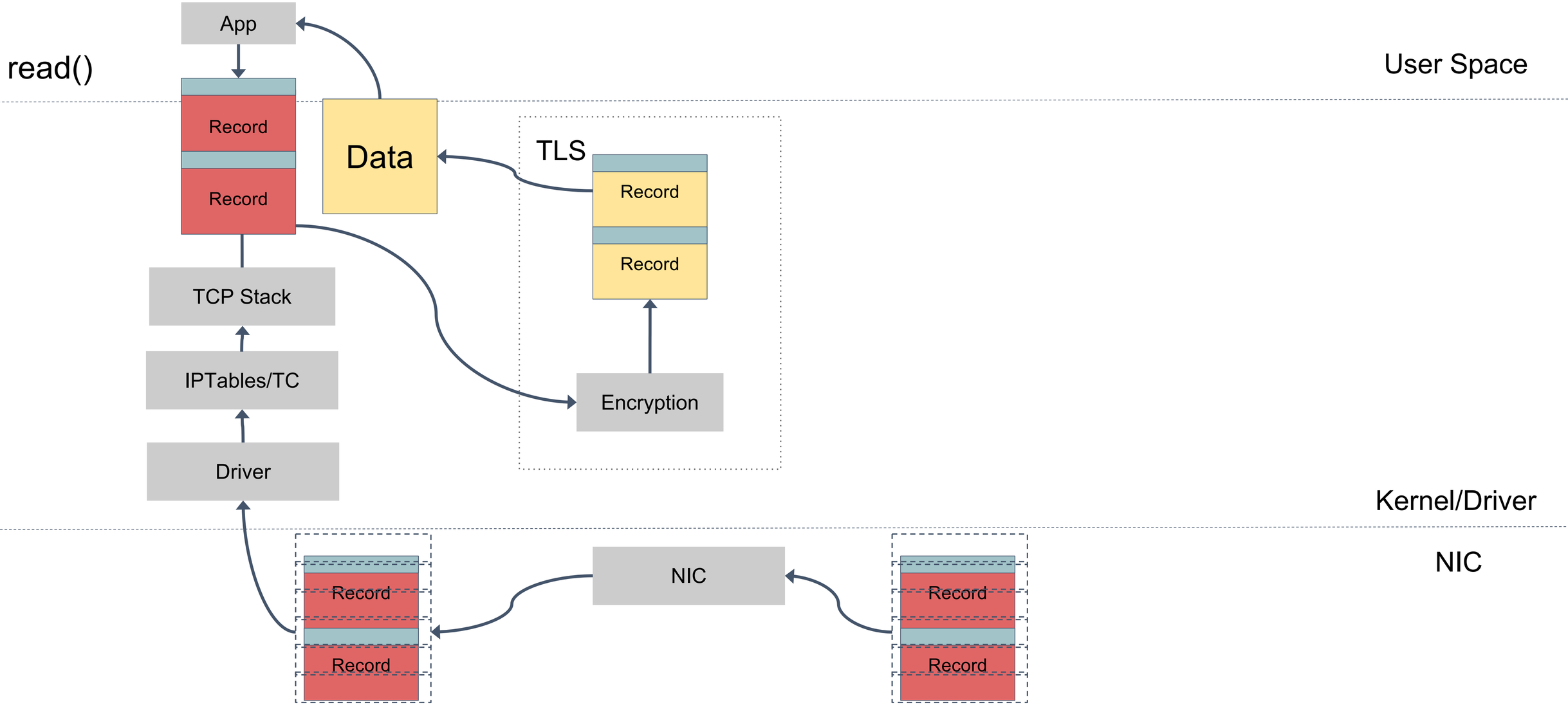
Kernel Stack: TX 50,000 ft Non-offloaded



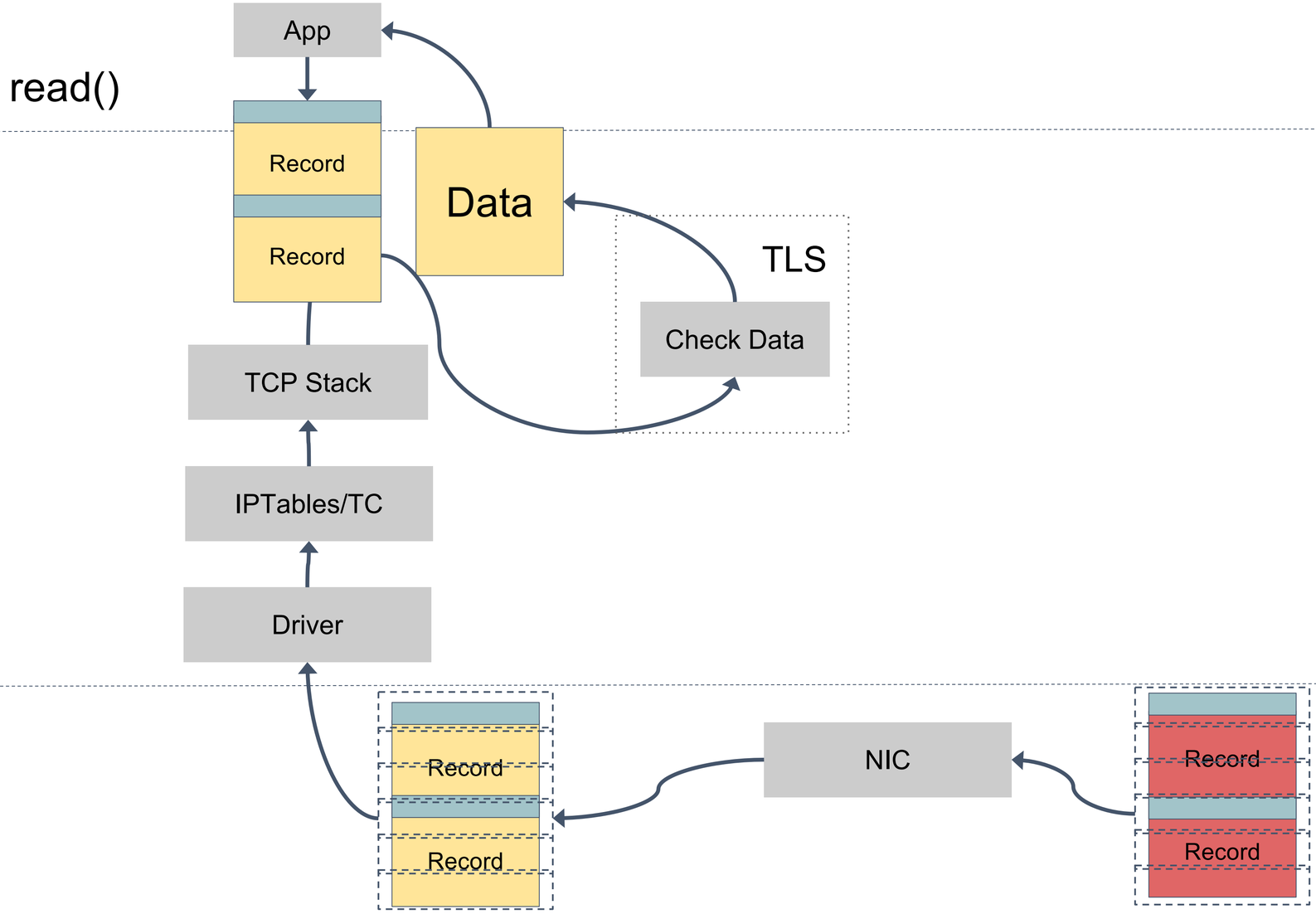
Kernel Stack: TX 50,000 ft Offloaded



Kernel Stack: RX 50,000 ft Non-offloaded

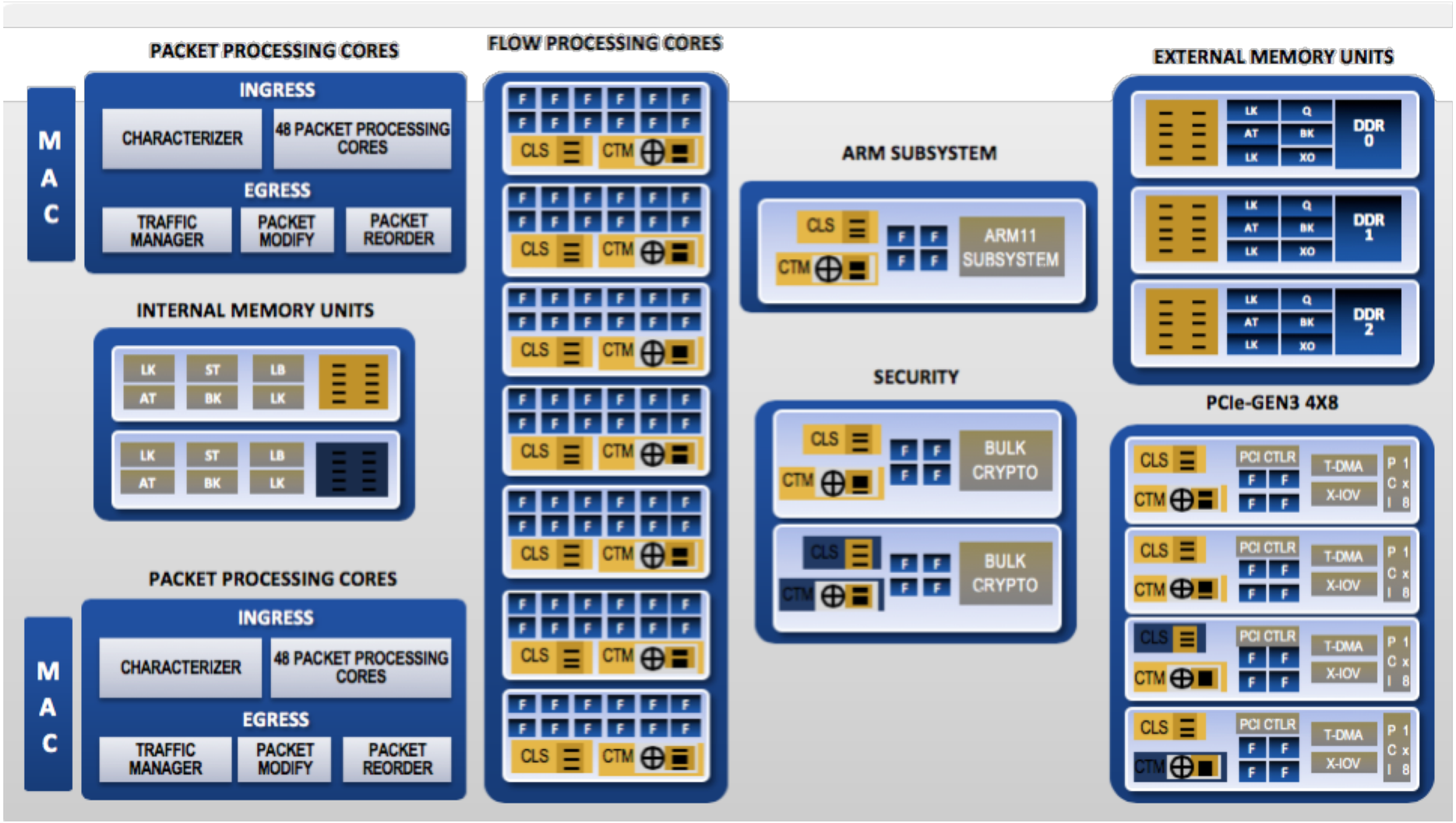


Kernel Stack: RX 50,000 ft Offloaded

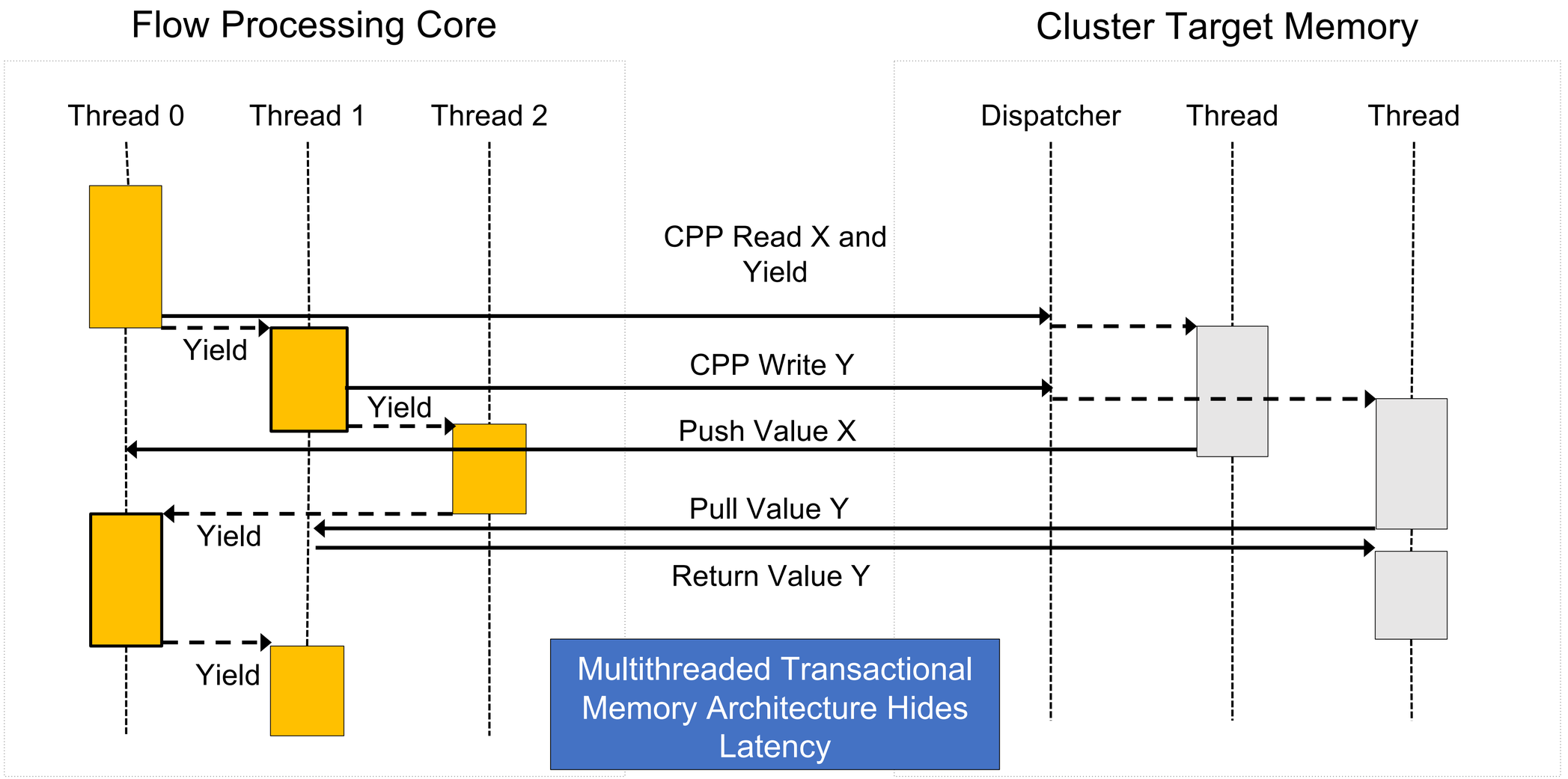


Offload-Silicon & FW

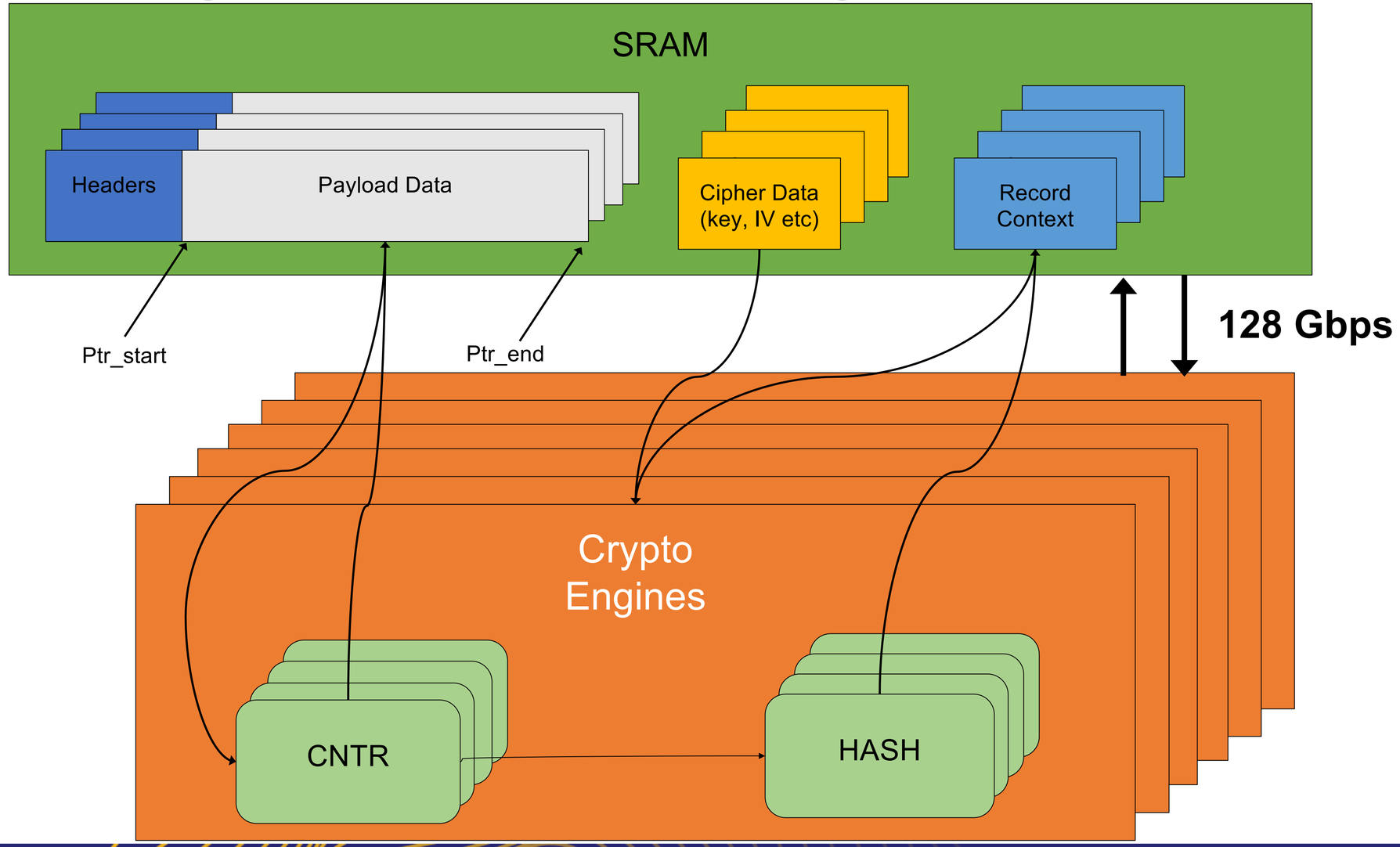
Background: Netronome NFP



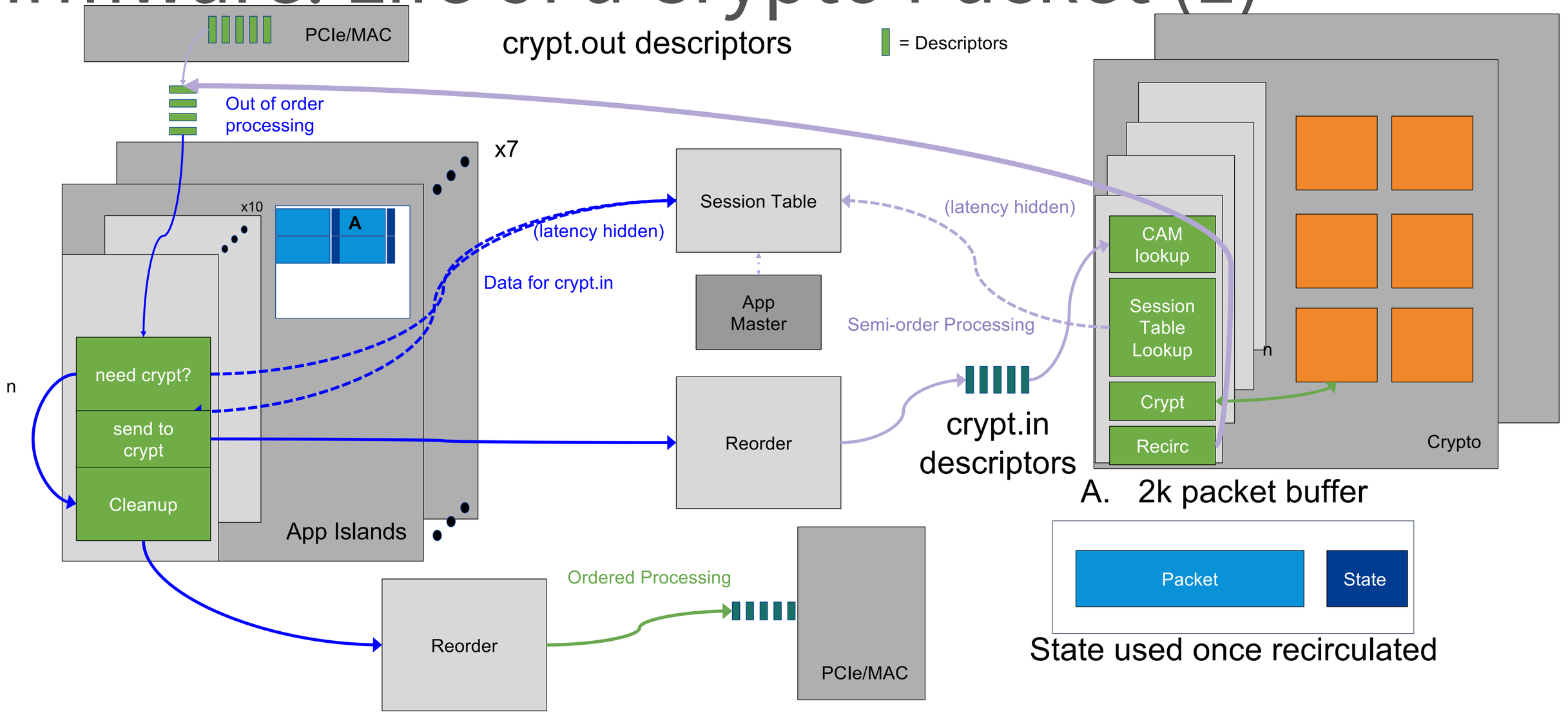
Memory Architecture



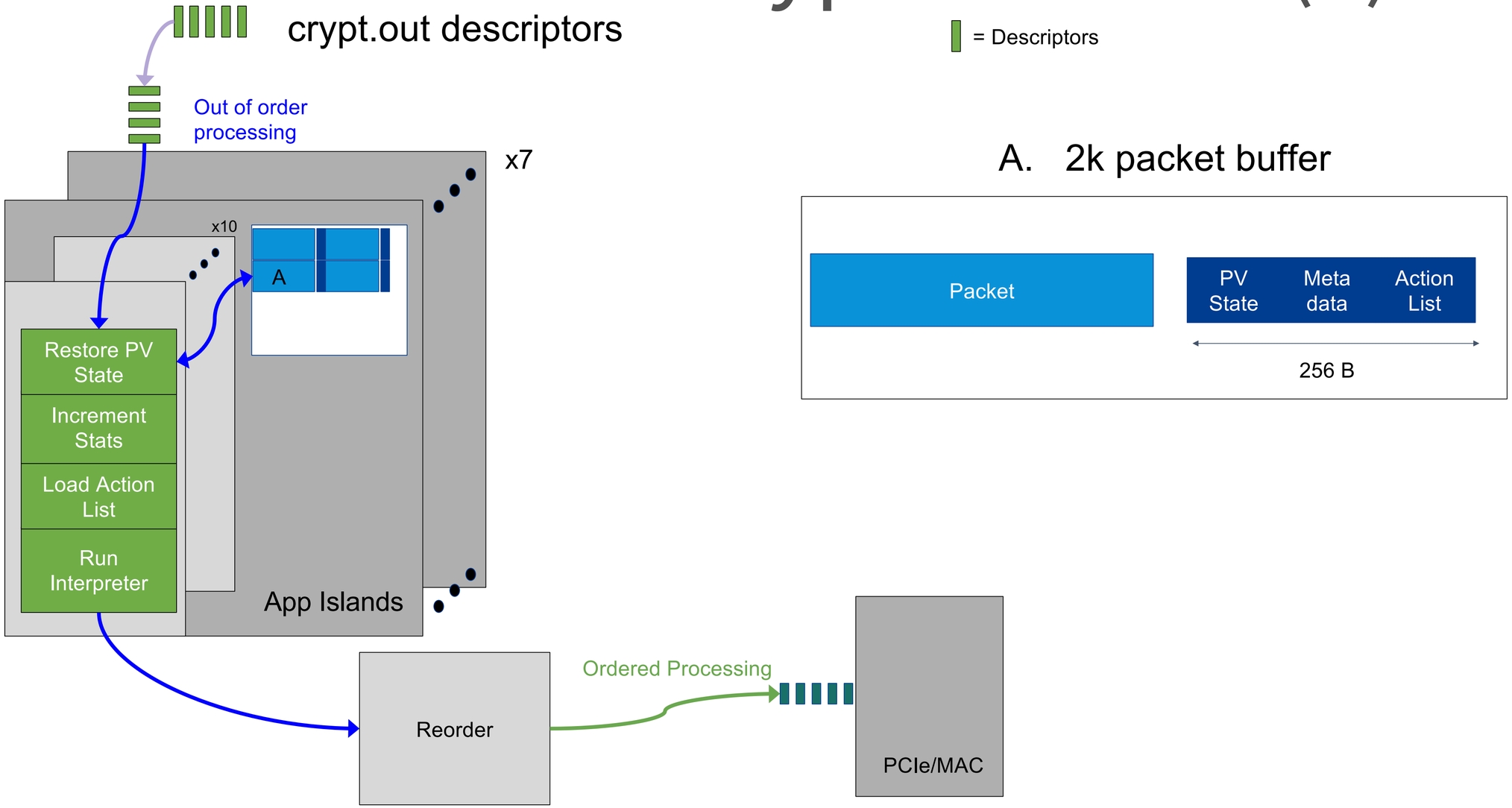
Crypto Engine Processing



Firmware: Life of a Crypto Packet (1)



Firmware: Life of a Crypto Packet (2)



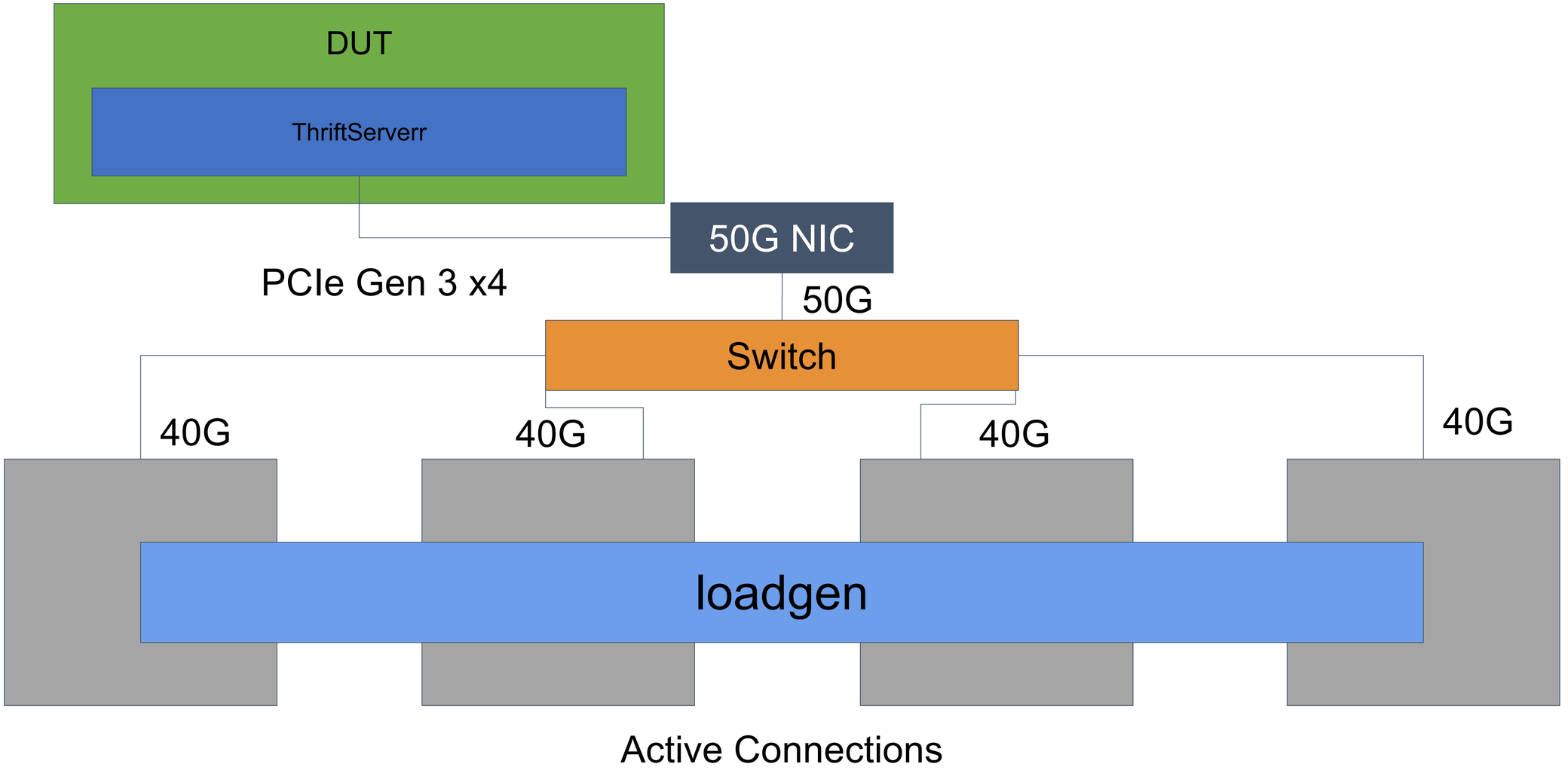
A. 2k packet buffer

Firmware: Open Source

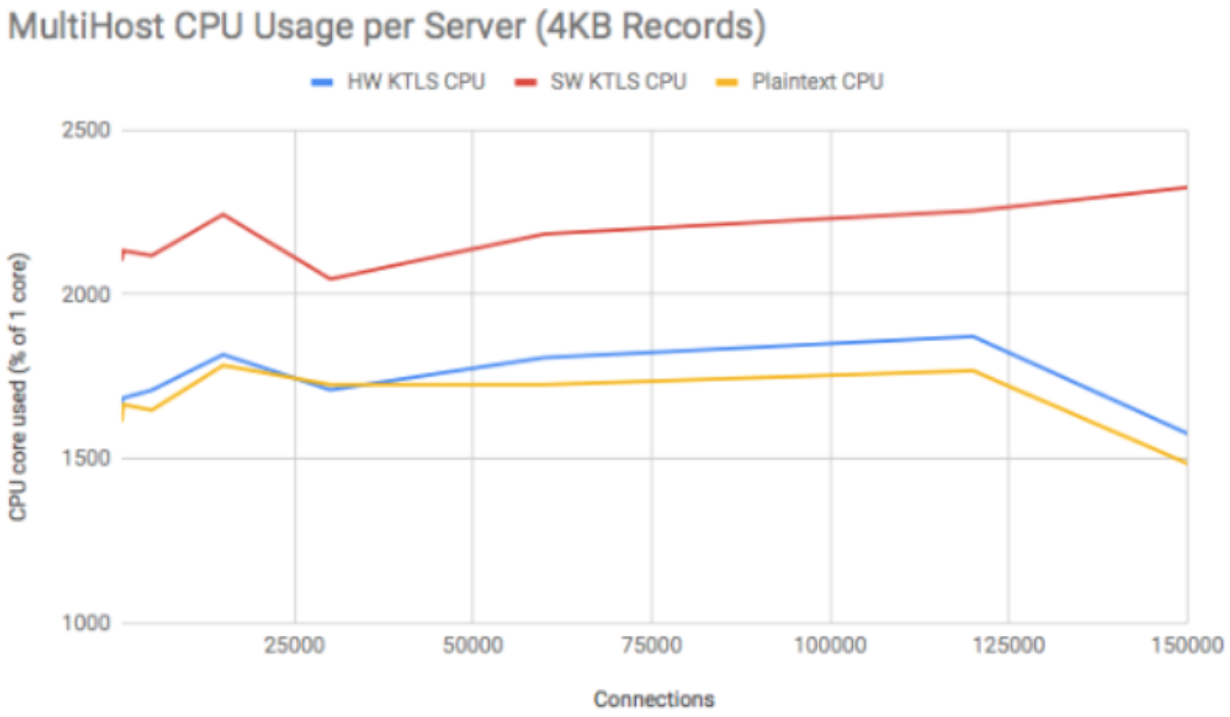
- We have open source our standard NIC FW
- Looking to incorporate this work relatively soon
- Allows others to contribute to our FW
- Allows customers the ability to see exactly what is happening
- See call to action slide for link!

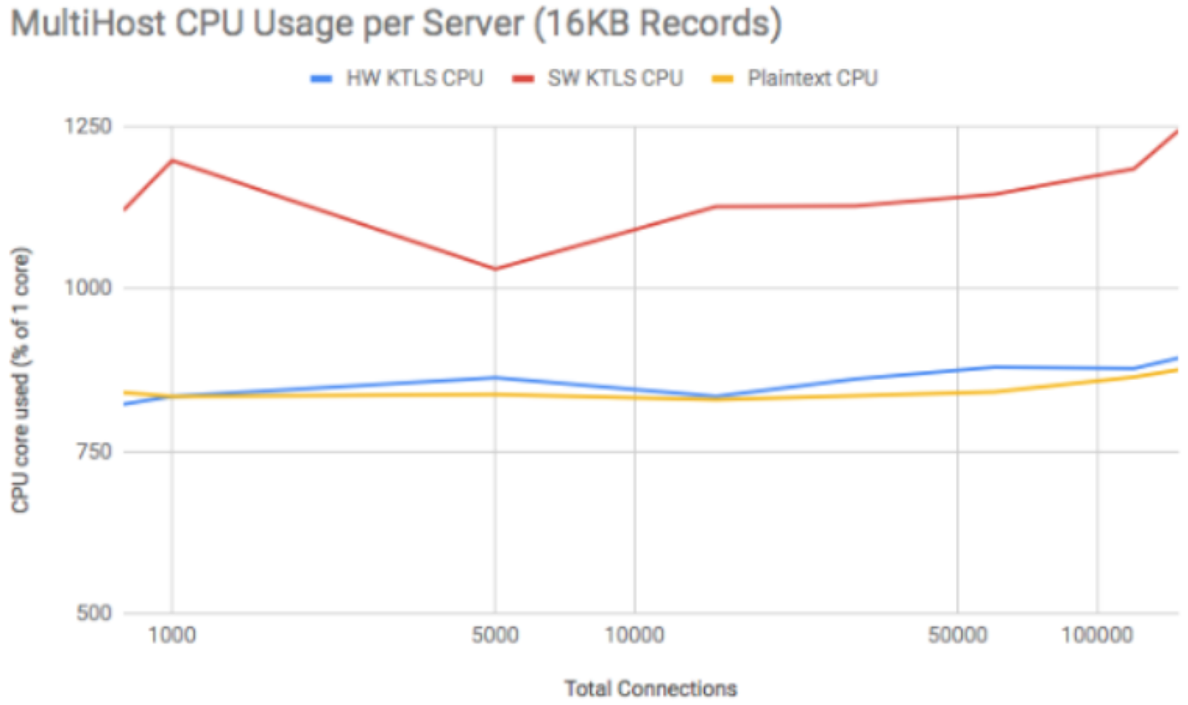
Benchmarks

Testing: Methodology



Benchmarks (at ~50 Gbps Line Rate)





Putting it all together

Summary

- TLS Offload returns up to over 90% of the CPU workload associated with crypto
  - Total CPU saved is related to the size of the records
  - The larger the records the more CPU saved:
- NFP Based SmartNICs offload this at low cost and power
- Done through the use of a domain specific architecture
  - This lends itself well to handling TLS based crypto for 100,000s of connections

Product Info

Agilio-CX 50G OCP Mezz 2.0 NIC

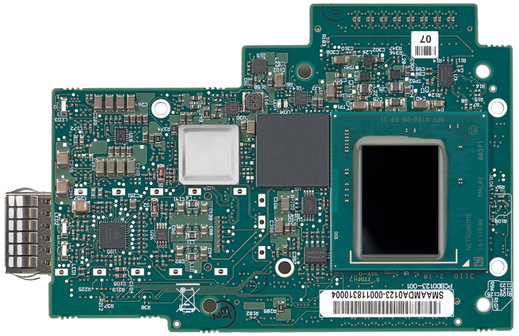
https://www.opencompute.org/wiki/Server/Mezz#Specifications_and_Designs

New Mezz v2 Type 5 Spec

http://files.opencompute.org/oc/public.php?service=files&t=5ad90059827e13e0273ce1446393225e

Work in Progress

- CLA signed for contributing Design Files
- Working on OCP Accepted™ product recognition



Call to Action

Netdev: netdev@vger.kernel.org

Open NIC FW: https://github.com/Netronome/nic-firmware

Open-NFP: open-nfp@googlegroups.com

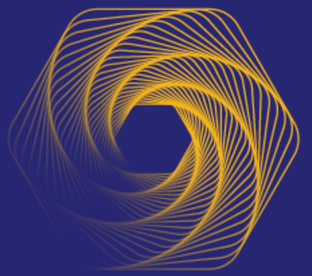
OCP Mezz: opencompute-mezz-card@lists.opencompute.org

Where to buy: https://www.netronome.com/products/agilio-cx/

Project Wiki with latest specification : http://www.opencompute.org/wiki/Server/Mezz

TLS Spec: http://lists.opencompute.org/mailman/listinfo/opencompute-mezz-card

KTLS kernel docs: https://www.kernel.org/doc/html/latest/networking/tls-offload.html

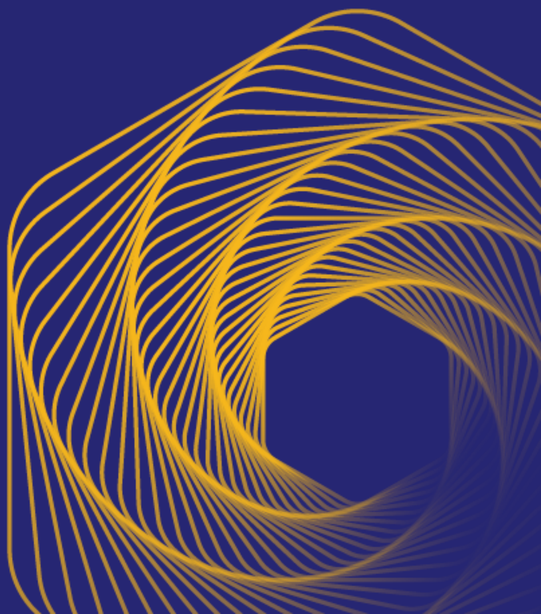


OCP REGIONAL SUMMIT

Open. Together.

OCP Regional Summit

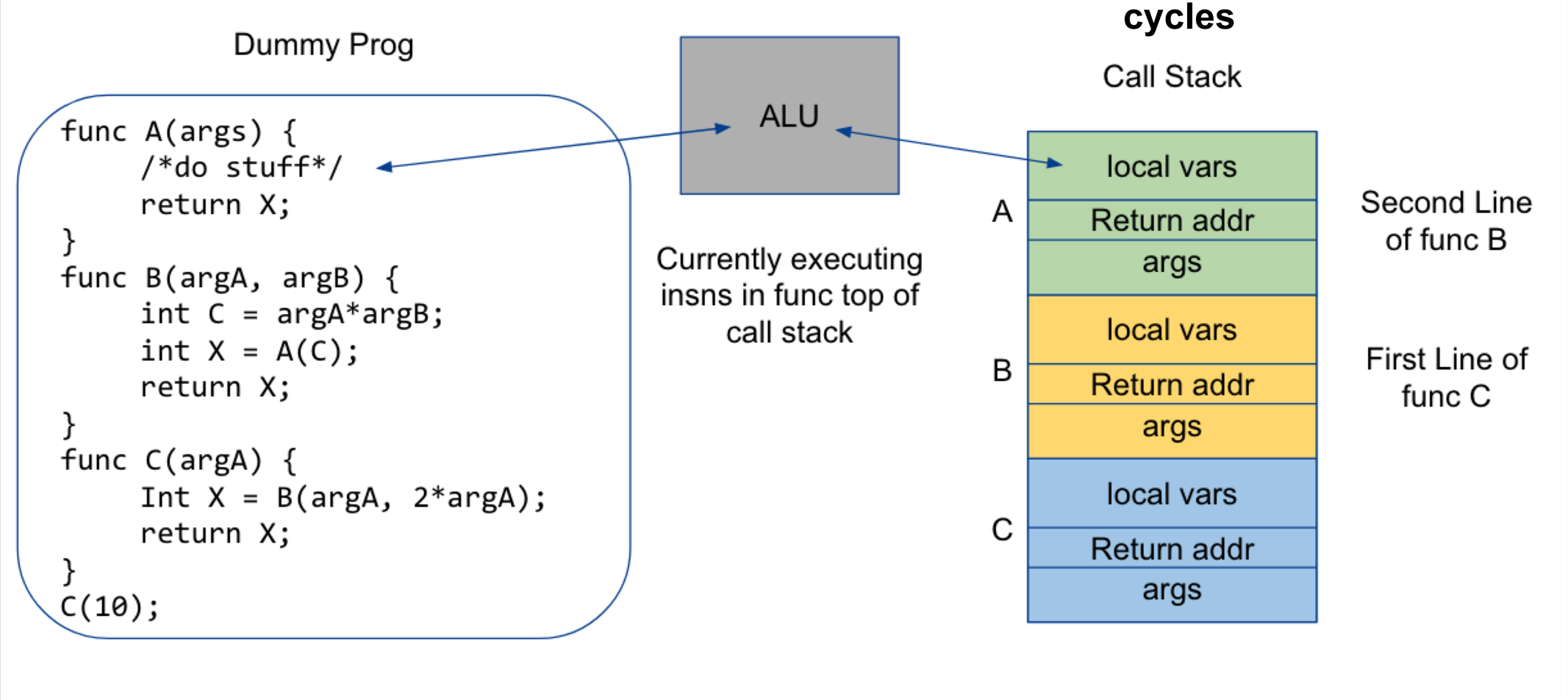
26–27, September, 2019



Backup:
Performance Analysis

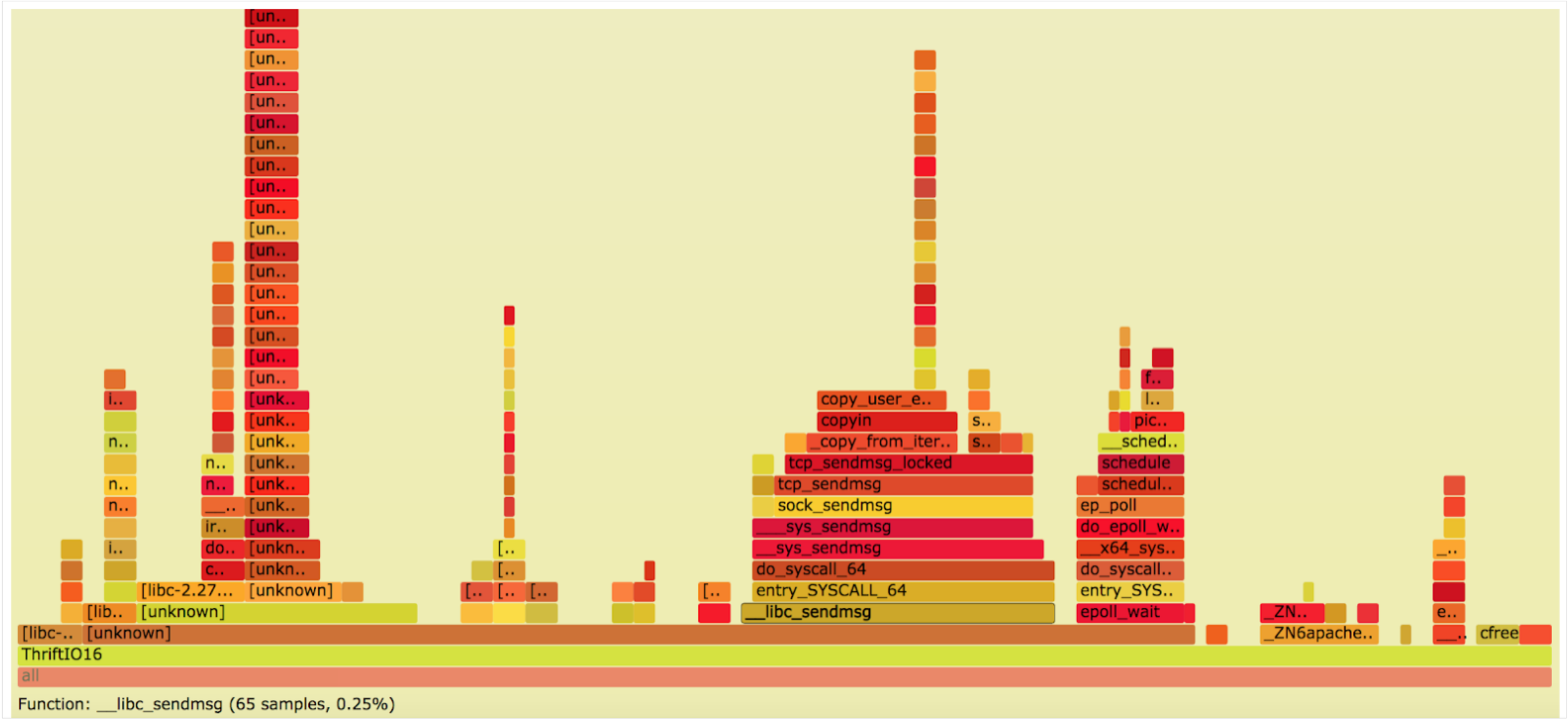
Performance Analysis: Flame Graphs

Flame Graphs are a histogram of call stack state



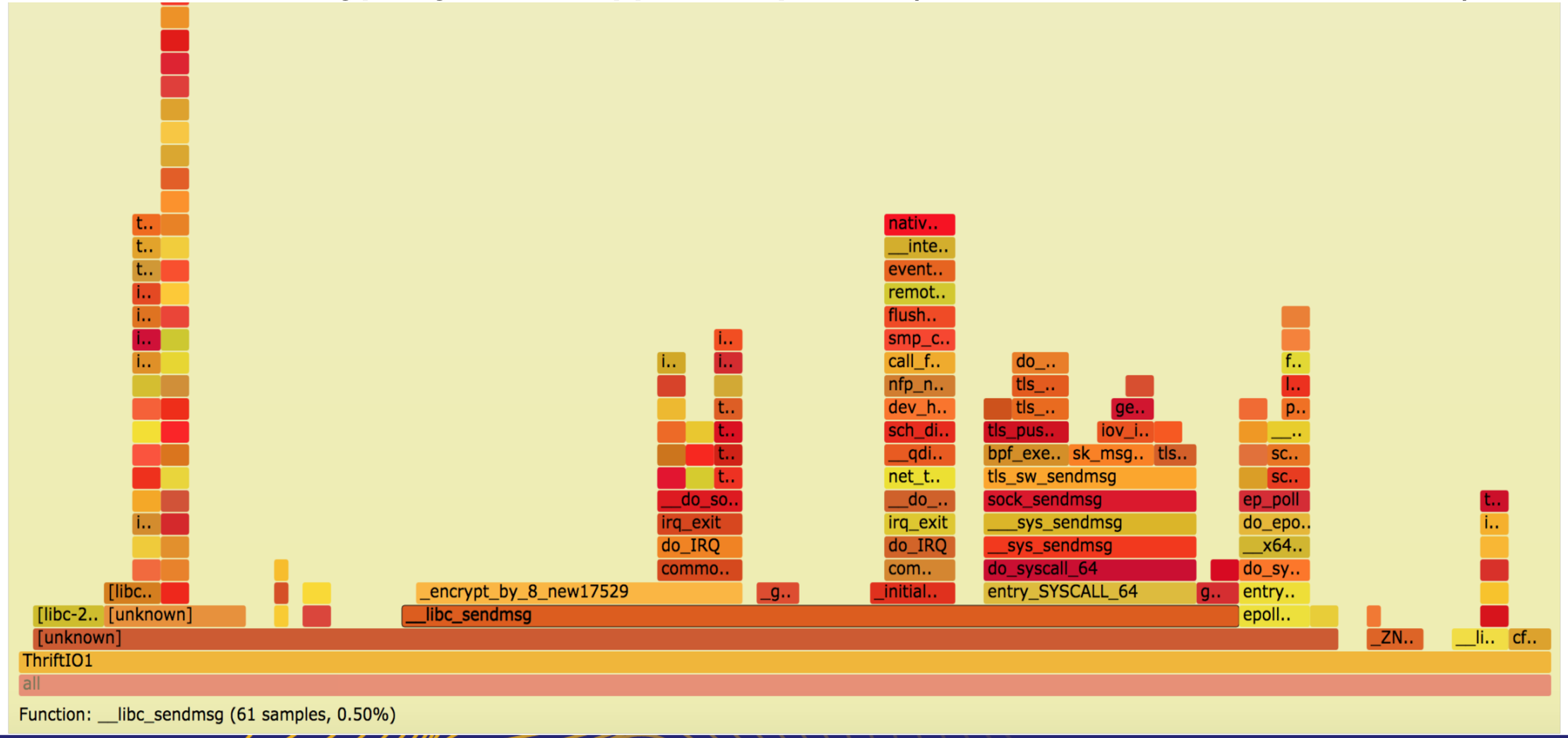
Performance Analysis: Cycles

Plaintext-cycles used by TCP stack (low connection count case-1000)



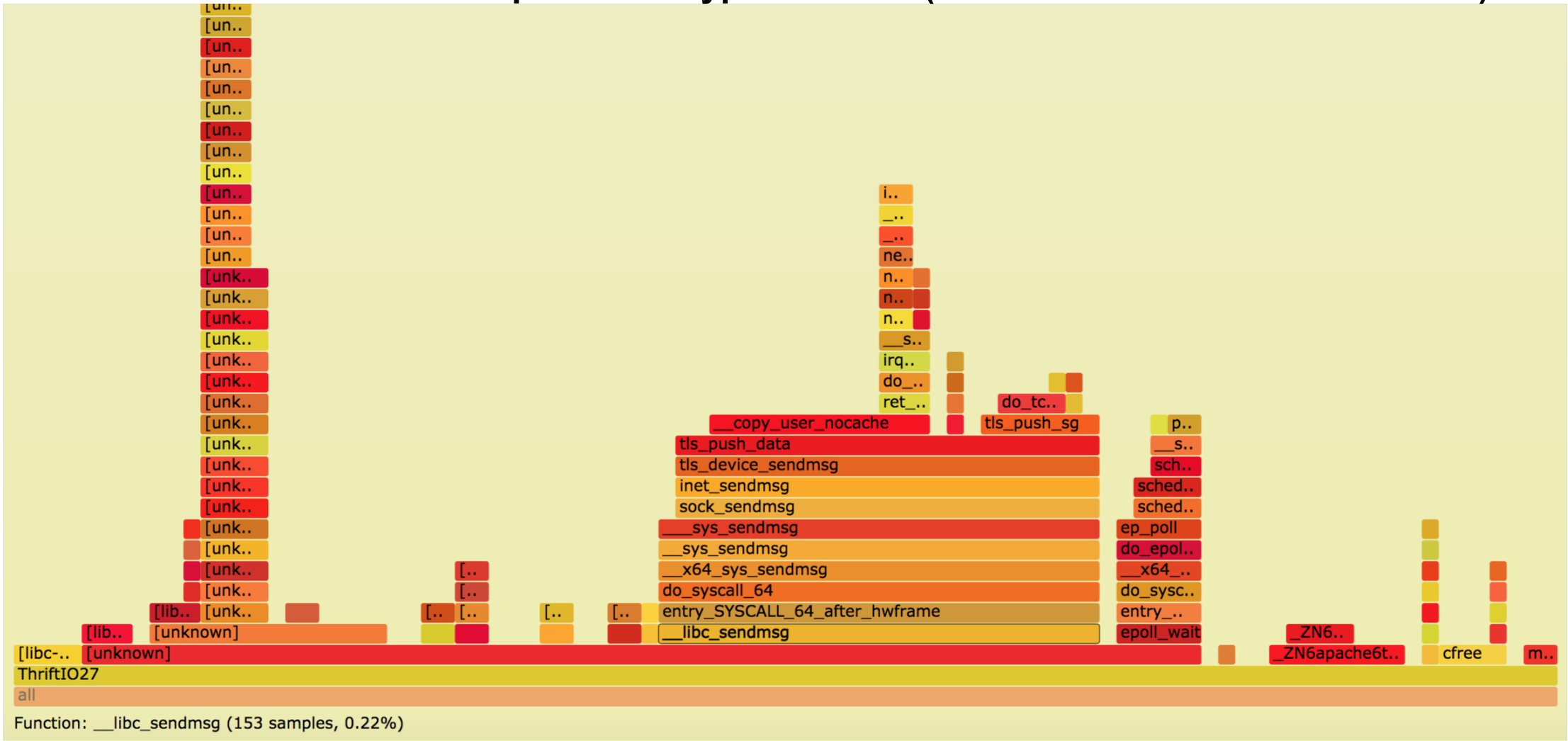
Performance Analysis: Cycles

SW KTLS-encrypt_by_8_new appears expensive(low connection count case-1000)



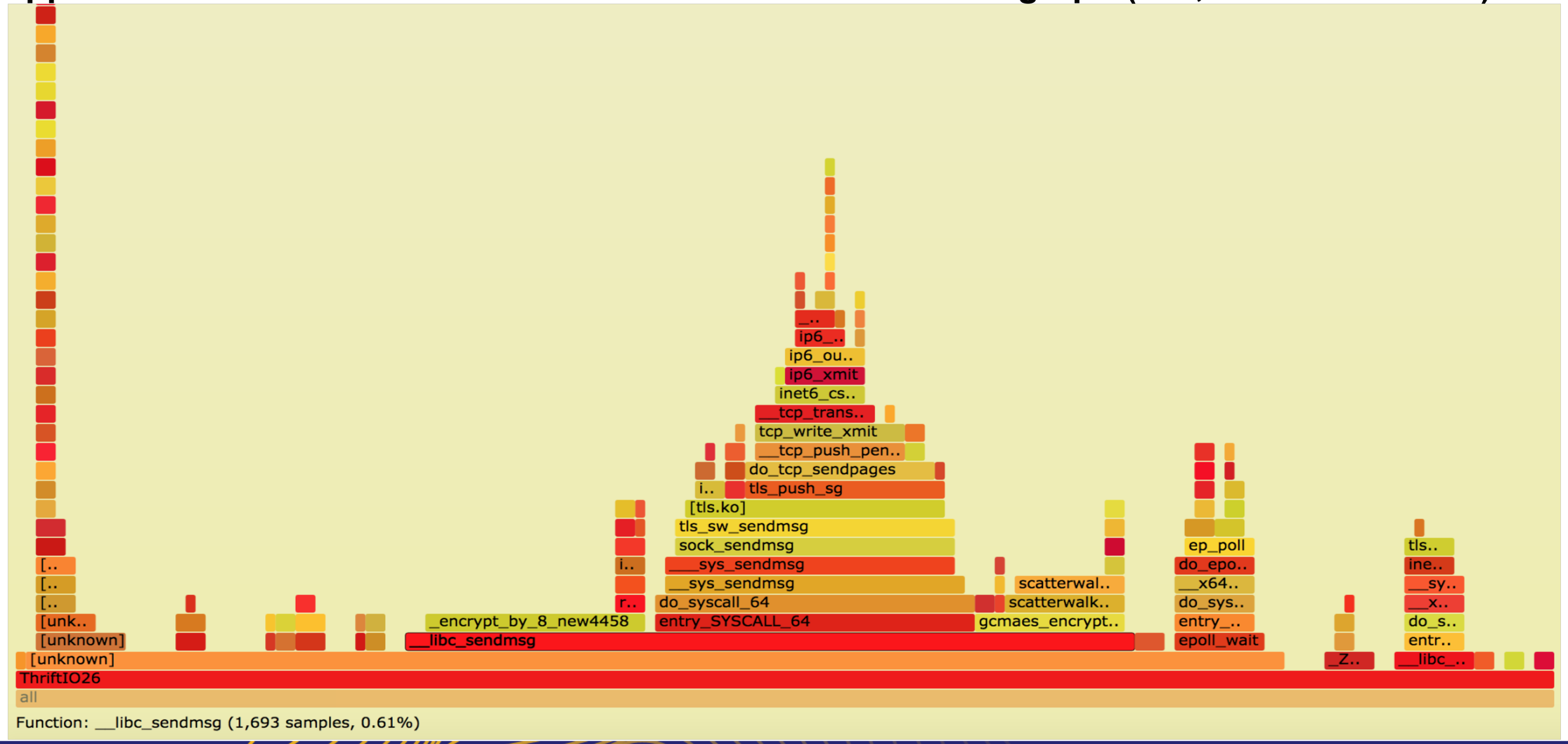
Performance Analysis: Cycles

HW KTLS-removes the expense of crypto on host (low connection count case-1000)



Performance Analysis: Effect of Connections

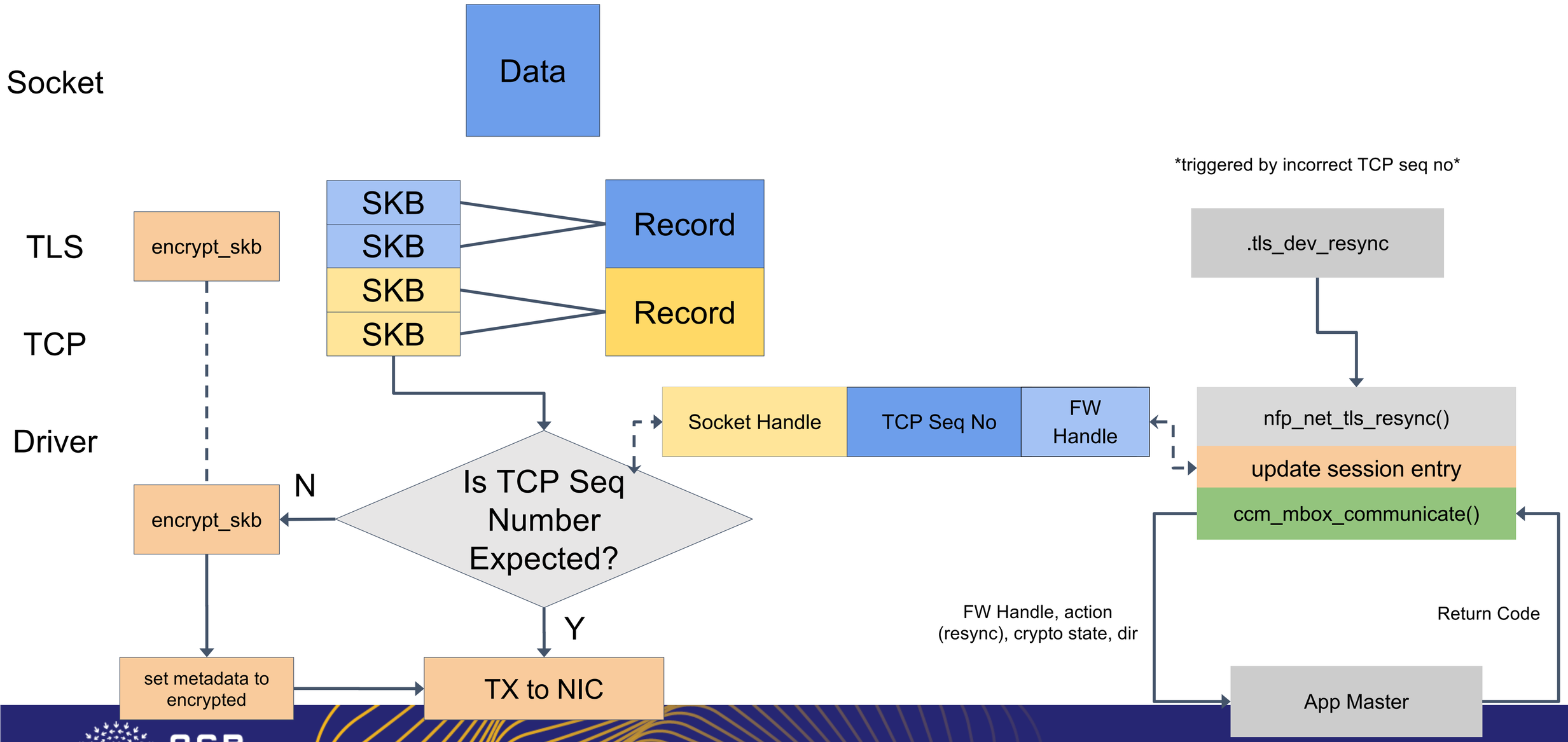
Appears to be due to different effect-see scatterwalk in flamegraph (150,000 connections)



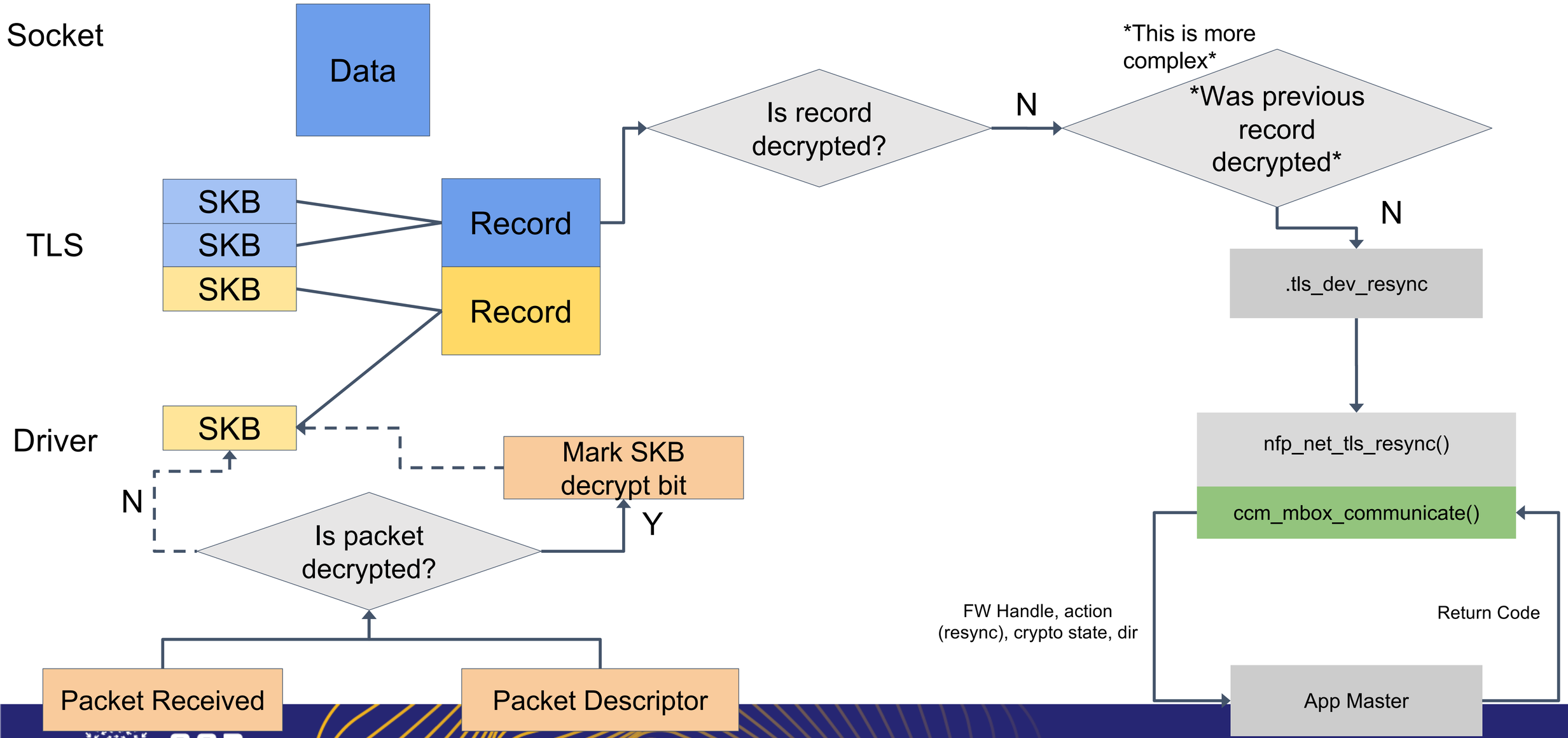
Performance Analysis: Cache Misses

- Potentially some correlation between cache misses and performance
  - ThriftServer:
    - 91% LLC hits for plaintext, HW KTLS
    - 75% for SW KTLS
  - Loadgen (appears more memory heavy)
    - 86% Plaintext, HW KTLS, ~25% for SW KTLS
- Loadgen also has 30% larger performance delta
- 50% of the misses appear to be stores sourced from encrypt_by_8_new
  - This is an optimized macro in linux/arch/x86/crypto/aesni-intel_avx-x86_64.S
- Further investigation required

Bulk Encryption: TX Path

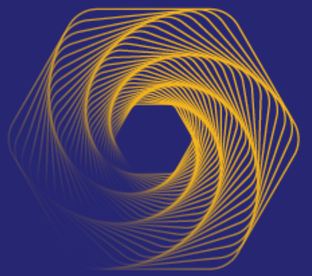


Bulk Encryption: RX Path



Moving to TLS 1.3

- Modify add mechanism to add the TLS type
  - No way of telling from the packets-masquerade as 1.2
- New encryption instruction lists for the TLS 1.3
- Modify the record reassembly as nonce cases don’t apply anymore
- This work would be post beta-testing success in current schedule

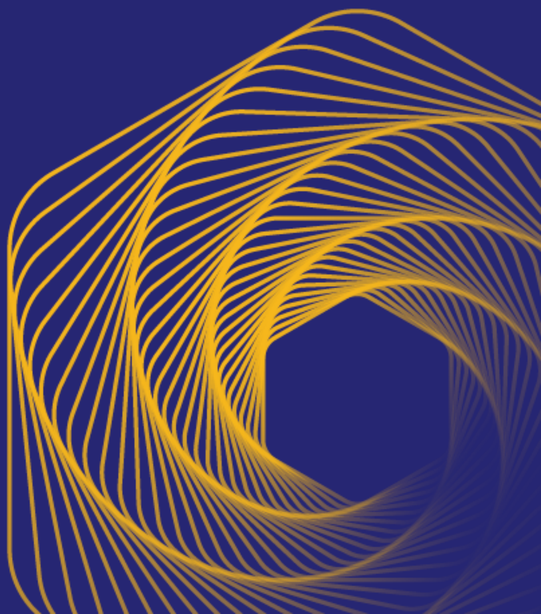


OCP REGIONAL SUMMIT

Open. Together.

OCP Regional Summit

26–27, September, 2019



Please use the appropriate icon representing the Project Group

SECURITY

SERVER

STORAGE

NETWORKING

RACK & POWER

MANAGEMENT

HPC

TELCO

DATA CENTER FACILITIES

OPEN SYSTEMS FIRMWARE

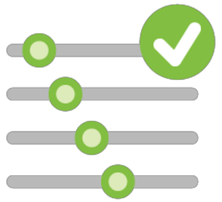
The following project group logos are missing: OpenEdge, OpenRMC, ACS. If you need one of these, contact Archna@opencompute.org

Please use the appropriate icon representing your type of contribution

Specifications

Reference Architecture

Embedded Software



Tested Configurations

Case Studies

White Papers

Design Files

Product Recognition

Facility Recognition

Workshops Summits

Testimonials Seminars

Videos