



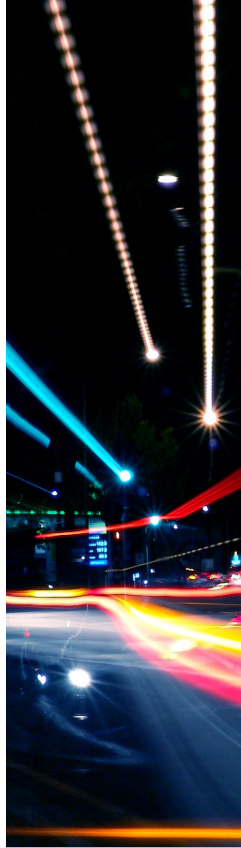
Programmable Hardware for UPF

Instrumentin g Latency

Dr Ben Lewis

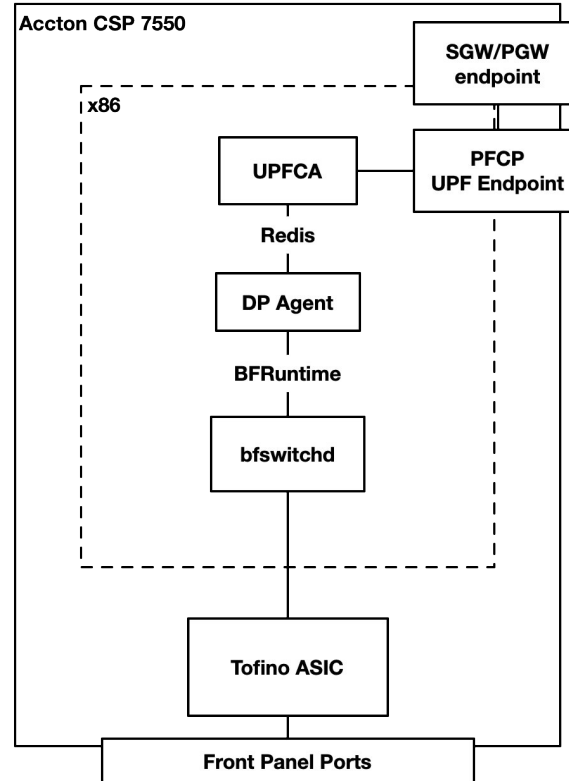
Motivations

- Evaluating the performance of our Tofino-based UPF using resources available in the i14y lab
 - Tofino programmable hardware has variable latency, depending on pipeline complexity



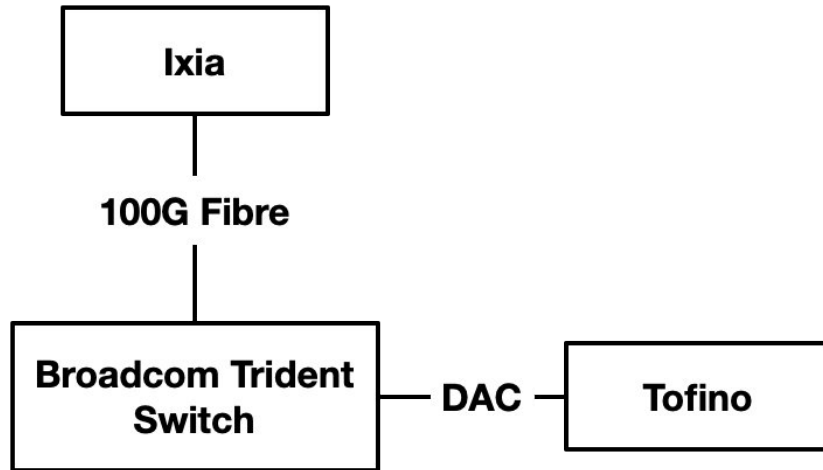
System Architecture

- Slicing
- Usage Reporting
- Coexisting 4G and 5G modes
 - 4G SGWu and PGWu endpoints
 - 5G UPF endpoint
- Golang-based control plane and PFCP implementation
 - (<https://github.com/bisdn/pfcpcore>)
- Protobuf-based interaction with Tofino drivers
- Redis used between upper and lower control plane layers
 - Separates the platform-specific processing
- P4 pipeline supporting >100k user sessions



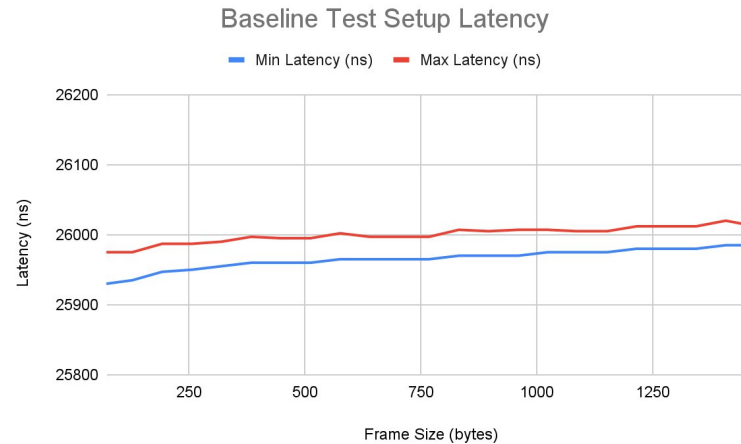
Evaluation Setup

- Tofino lacks support for some QSFP modules
- Agema AG5648 (with Broadcom Trident) used to bridge the connection



Baseline Latency

- We can first use the Ixia to establish the delay introduced by the fibre and AG5648 switch
- Ixia has a preset for RFC2544 (Benchmarking Methodology for Network Interconnect Devices)
- Latency of our test setup
Minimum: 25930ns
Maximum: 26012ns
Average variation: 1.1ns



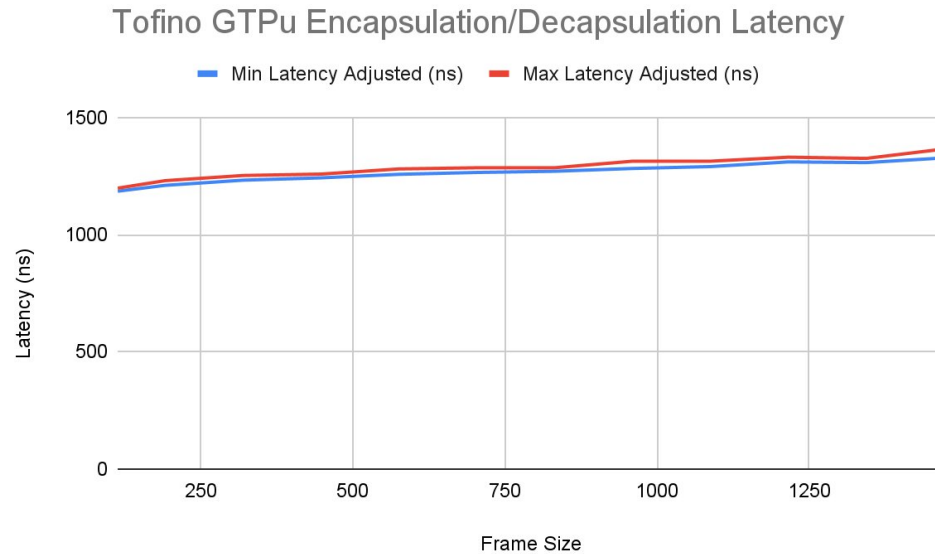
Tofino UPF Test Scenario

- Using the Tofino UPF alters the test size slightly, as the minimum frame size is increased due to encapsulation requirements.
- Evaluation of the GTPu encapsulation/decapsulation
 - Control plane is out of scope in this test
- 20k sessions installed into the data-plane
 - One upstream and one downstream entry per session
 - Aligned with Ixia configurations

Results

Minimum latency: 1185ns
Maximum latency: 1365ns

Average variation for all frame sizes: 1.1ns



Conclusions

- Using equipment available in the i14y lab, we can effectively measure:
 - Performance of our DUT
 - Performance of the connecting infrastructure
- Tofino can offer performant dataplane encapsulation and decapsulation
 - Pipeline configurations can alter the latency profile

