A Hybrid FPGA-P4 Architecture for 400G Line-Rate Packet Processing

Yu-Kuen Lai, Yu-Chen Cheng, Yi-Siang Lin,
Miura Hideyoshi*, Chong-Hao Kuo
Department of Electrical Enginering
Chung-Yuan Christian University, Chungli, Taiwan
*Graduate School of Interdisciplinary Science and
Engineering in Health Systems
Okayama University, Japan

Jim Hao Chen, Fei Yeh, Joe Mambretti
International Center for Advanced Internet Research
Northwestern University, USA

Abstract

This work presents a programmable, FPGA-accelerated framework for real-time packet header processing in high-speed networks. By aggregating up to four packet headers and leveraging stream-based second frequency moment estimation with parallel count sketches on an AMD U200 FPGA and Tofino2 P4 switch, the system achieves efficient and scalable traffic analysis at 400 Gbps. A preliminary evaluation using MAWI and CAIDA DDoS traces demonstrates accurate estimation performance compared to exact calculations. The design reduces processing overhead while maintaining line-rate throughput, making it suitable for next-generation monitoring and anomaly detection.

1 Introduction

Traffic analysis is a foundational component of modern network systems. As data-intensive services proliferate and cloud computing becomes increasingly central to global infrastructure, the volume and velocity of network traffic have surged. In particular, core network links now operate at transmission speeds of 400 Gbps and beyond, pushing the limits of conventional traffic monitoring and analysis tools. These extreme data rates necessitate the development of specialized systems capable of processing traffic in real time without incurring excessive latency or sacrificing accuracy [1].

Building on the insights and framework proposed in our previous work in NRE of SC24, this study introduces a highly flexible real-time packet header processing architecture. The design leverages the principles of Software-Defined Networking (SDN), particularly those embodied in programmable SDN switches, to enable the dynamic configuration of packet-handling behaviors. By using the P4 language, the system supports the aggregation of header information from up to n consecutive packets (where n ranges from 1 to 4). This feature significantly reduces the processing overhead in the early stages of analysis.

Central to the proposed framework is the use of Field-Programmable Gate Arrays (FPGAs), whose inherent parallelism and reconfigurability make them ideal for high-speed data path applications. Through hardware acceleration, the system not only meets stringent throughput requirements but also offers scalable performance across varying network conditions. This combination of programmability, efficiency, and hardware-accelerated speed makes the architecture well-suited for deployment in next-generation network monitoring and security infrastructures.

2 Goal and Impact

The primary goal of this research is to design, implement, and validate a flexible, low-overhead architecture for real-time second frequency moment(F2) estimation. By integrating an FPGA with a P4 programmable packet processing framework, we aim to create a solution that can be dynamically tailored to specific measurement tasks, overcoming the processing bottlenecks of traditional methods.

This research will provide the network security community with a tested, high-performance tool for real-time traffic analysis. The immediate impact is a significant reduction in processing overhead, making high-speed F2 estimation practical on production networks. The broader, long-term impact is enabling the development of advanced on-device applications, particularly in real-time anomaly detection, which is critical for securing high-speed networks against emerging threats. Future work will focus on expanding this architecture to support even more advanced analytics.

3 FPGA Design

To support high-throughput traffic monitoring in real-time, the proposed architecture integrates an FPGA-based hardware accelerator into one of the monitoring nodes.

The design, based on the NetFPGA-Plus framework [3], demonstrates a stream-based implementation of the second frequency moment estimation algorithm, utilizing multiple parallel count sketch instances. By aggregating headers from multiple packets (up to four), the input data rate is significantly reduced while retaining essential statistical information for flow-level analysis.

As illustrate in Figure 1, upon receiving aggregated headers, the FPGA performs parallel parsing and simultaneous updates across all sketch pipelines. This parallelism is crucial in maintaining throughput scalability, ensuring the system

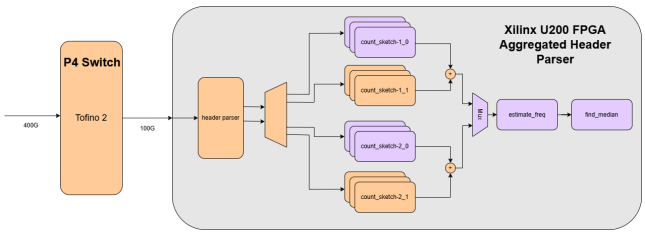

Figure 1. The implementation of the second frequency moment estimation algorithm based on multiple parallel count sketch instances.

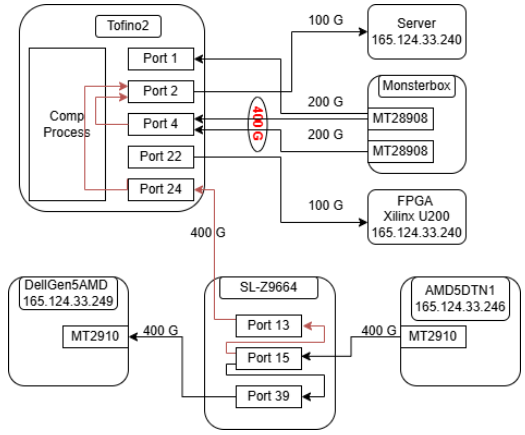

Figure 2. The iCair P4/FPGA programmable real-time packet header processing testbed.

can handle the full 400 Gbps of input traffic without introducing performance bottlenecks.

4 System Testbed

The network testbed, as illustrated in Figure 2, is designed to evaluate high-speed traffic monitoring using a programmable data plane and FPGA-based acceleration. At the core of the testbed is a Tofino2 switch, which processes selected ingress traffic internally and forwards it to the remote nodes.

Test traffic, reaching rates of up to 400 Gbps, is generated by high-performance servers: the Monsterbox, equipped with two 200 Gbps MT28908 network interfaces, and AMD5DTN1, featuring a single 400 Gbps MT2910 interface.

5 Preliminary Test

The effectiveness of the proposed design was evaluated using real-world network traffic traces. Specifically, a one-hour packet capture from the MAWI Working Group Traffic Archive (MAWI 2019 DITL Trace) [2] was utilized as background traffic. To simulate an attack scenario, this background traffic was merged with the CAIDA DDoS 2007 attack trace. Figure 3 illustrates a comparison between the FPGA-based estimation (lower panel) of the second frequency moment—computed using destination IP addresses—and the exact

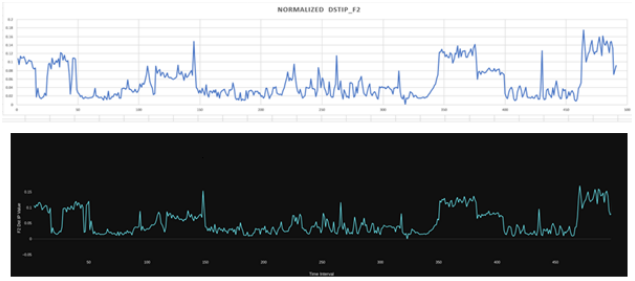

Figure 3. The second frequency moment computed based on destination IP addresses shows high accuracy in the FPGA-based estimation, using aggregated received headers (lower panel) compared to the exact calculation (upper panel).

values obtained through software-based calculation (upper panel). This comparison highlights the FPGA's capability to approximate traffic statistics accurately in real time while handling high-throughput workloads.

6 Conclusion and Future Work

This work-in-progress presents a stream-based implementation of second frequency moment estimation on an FPGA integrated within a P4 programmable real-time packet header processing framework. The proposed architecture enables the aggregation of packet headers across multiple packets, allowing it to be tailored to the demands of specific measurement tasks. This approach not only enhances the system's flexibility but also significantly reduces processing overhead, making it well-suited for high-speed traffic analysis. Moving forward, future work will focus on further exploring the architectural design space and refining the implementation to support advanced applications, particularly in the area of high-speed network anomaly detection.

References

[1] Sebastian GallenmÃŒller, Paul Emmerich, Florian Wohlfart, Daniel Raumer, and Georg Carle. 2015. Comparison of frameworks for high-performance packet IO. In *2015 ACM/IEEE Symposium on Architectures for Networking and Communications Systems (ANCS)*. 29–38. https://doi.org/10.1109/ANCS.2015.7110118

[2] Yu-Kuen Lai, Cheng-Lin Tsai, Cheng-Han Chuang, Xiu-Wen Ku, and Jim Hao Chen. 2022. Tabular Interpolation Approach Based on Stable Random Projection for Estimating Empirical Entropy of High-Speed Network Traffic. *IEEE Access* 10 (2022), 104934–104953. https://doi.org/10.1109/ACCESS.2022.3210336 Conference Name: IEEE Access.

[3] Tokusashi Yuta. 2021. NetFPGA/NetFPGA-PLUS. https://github.com/NetFPGA/NetFPGA-PLUS original-date: 2021-08-18T14:08:00Z.