中原大學

iCAIR

Real-Time Frequency Moment Estimation on FPGA: Applications in Anomaly Detection and Weibull Flow Length Parameterization

Yu-Kuen Lai, Hsiang-Lun Hua, Bo-Shun Huang
Department of Electrical Engineering
Chung-Yuan Christian University, Chungli, Taiwan

Jim Hao Chen, Joe Mambretti

International Center for Advanced Internet Research Northwestern University, USA

Presenter: Bo-Shun Huang

Department of Electrical Engineering Chung-Yuan Christian University, Chungli, Taiwan

電機工程學系

Frequency Moment

$$F_k = \sum_{i=1}^{n} (f_i)^k$$

- **A data stream** $\phi = (a_1, a_2, \ldots, a_n)$
- **Statistical moment of a frequency distribution**
- **A mathematical quantity that describes the characteristics of a probability distribution**
  - *n* represents the total number of distinct items
  - f_i is the frequency of each item a_i in the data stream
- F_0**: Total Distinct items**
- F_1**: Total Number of items**

Frequency Moment (2)

- $F_2 = \sum_{i=1}^{n} (f_i)^2$
  - known as the Gini’s homogeneity index
  - used to measure the variability and inequality in a frequency distribution
  - represent the degree of
    - spread-out
    - concentration
- **For $\mathrm{K} \geq 2$**
  - The degree of skewness of a given distribution
- **Frequency Moments can be used to gain insights into various unique features of traffic flows**

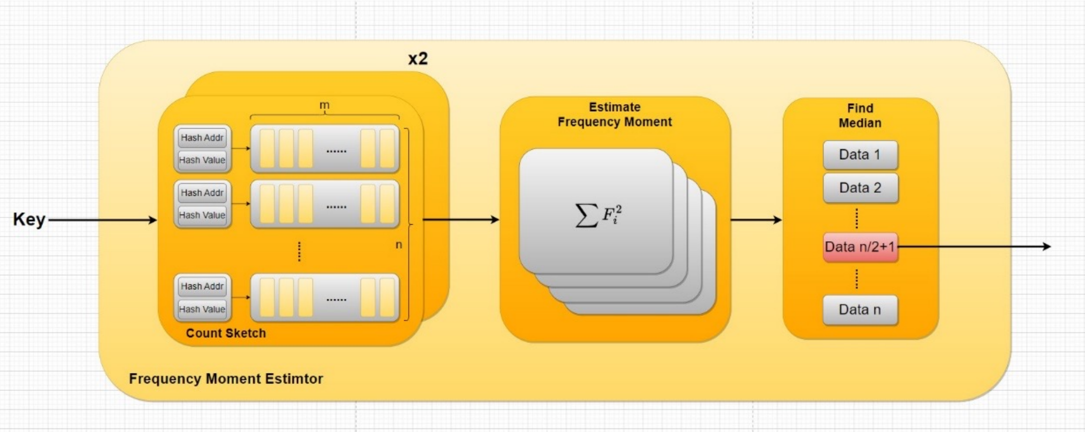
Challenges

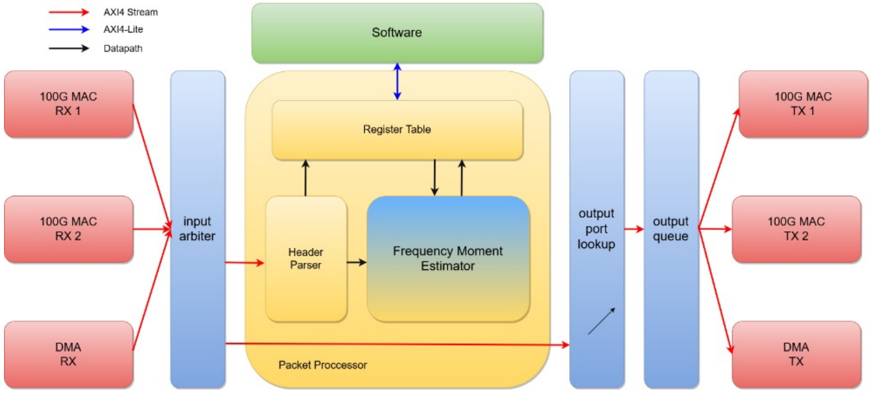
- **Online measuring of Frequency Moments on Internet traffic**
  - Attractive for many network applications
    - Anomaly detection, traffic analysis
- **How to process and compute statistics on data streams in real-time?**
- **Packet arrives at a rapid rate**
- **Key space**
  - IPv4 address of 32-bit
  - High distinct number of flows
- **Current Status**
  - Software-based
  - Off-line approaches

A. K. Marnerides, A. Schaeffer-Filho, and A. Mauthe, "Traffic Anomaly Diagnosis in Internet Backbone Networks,"
B. Computer. Networks, vol. 73, no. C, pp. 224–243, Nov. 2014.

Sketch-based Implementation

- ❑ **F_2 Estimation**
  - ○ AMS-Cormode Sketch base on Count-sketch [5]
- ❑ **F_0 Estimation**
  - ○ FM Sketch, Probabilistic Counting with Stochastic Averaging (PCSA) [6]
- ❑ **NetFPGA PLUS Framework**
- ❑ **Xilinx Alveo U200 FPGA**





[5] Graham Cormode and Marios Hadjieleftheriou. 2009. Finding the frequent items in streams of data. *Commun. ACM* 52, 10 (2009), 97–105. https://doi.org/10.1145/1562764.1562789

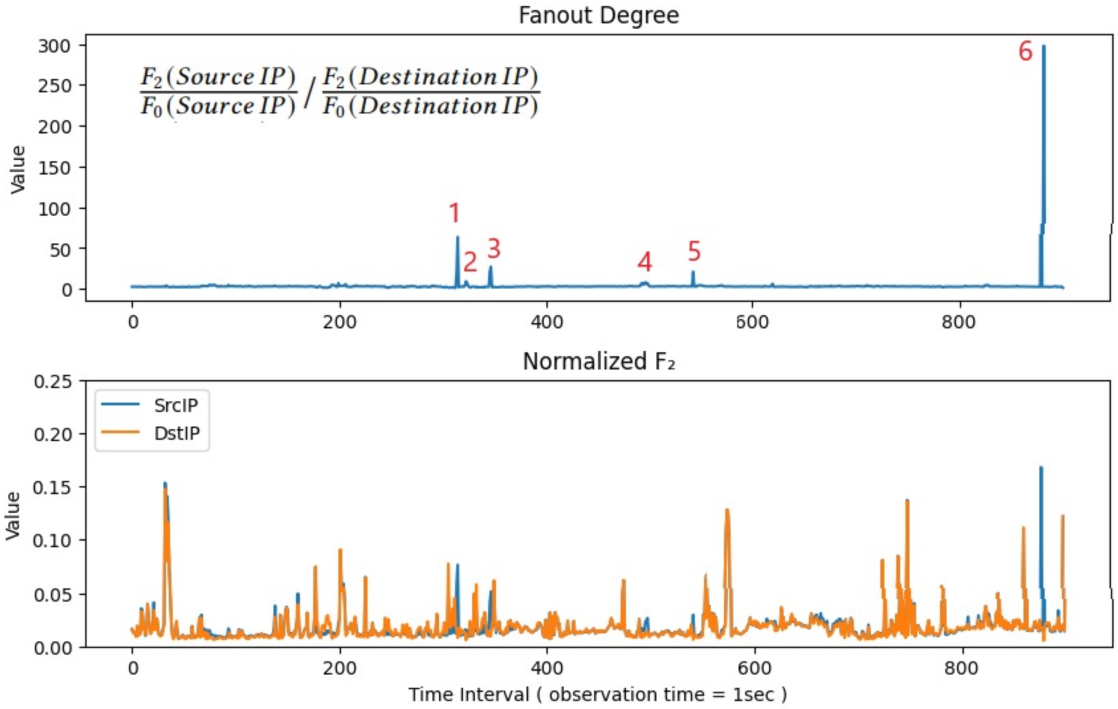
[6] Philippe Flajolet and G. Nigel Martin. 1985. Probabilistic counting algorithms for data base applications. *J. Comput. Syst. Sci.* 31, 2 (1985), 182–209. http://portal.acm.org/citation.cfm?id=5215

System Evaluation

- **Network Traffic Traces**
- **MAWI Samplepoint-F Trace**
  - scan anomaly detection
  - https://mawi.wide.ad.jp/mawi/samplepoint-F/2022/202201101400.html
- **DDoS**
  - CAIDA 2007 DDoS Trace (Attacking):
    - Four DDoS attacking traffic are selected from the CAIDA 2007 DDoS trace (to-victim).
      - 20070804_140436.pcap
      - 20070804_140936.pcap
      - 20070804_141436.pcap
      - 20070804_141936.pcap
  - MAWI DITL 2019 Trace (Background)
    - https://mawi.wide.ad.jp/mawi/ditl/ditl2019/201904091800.html

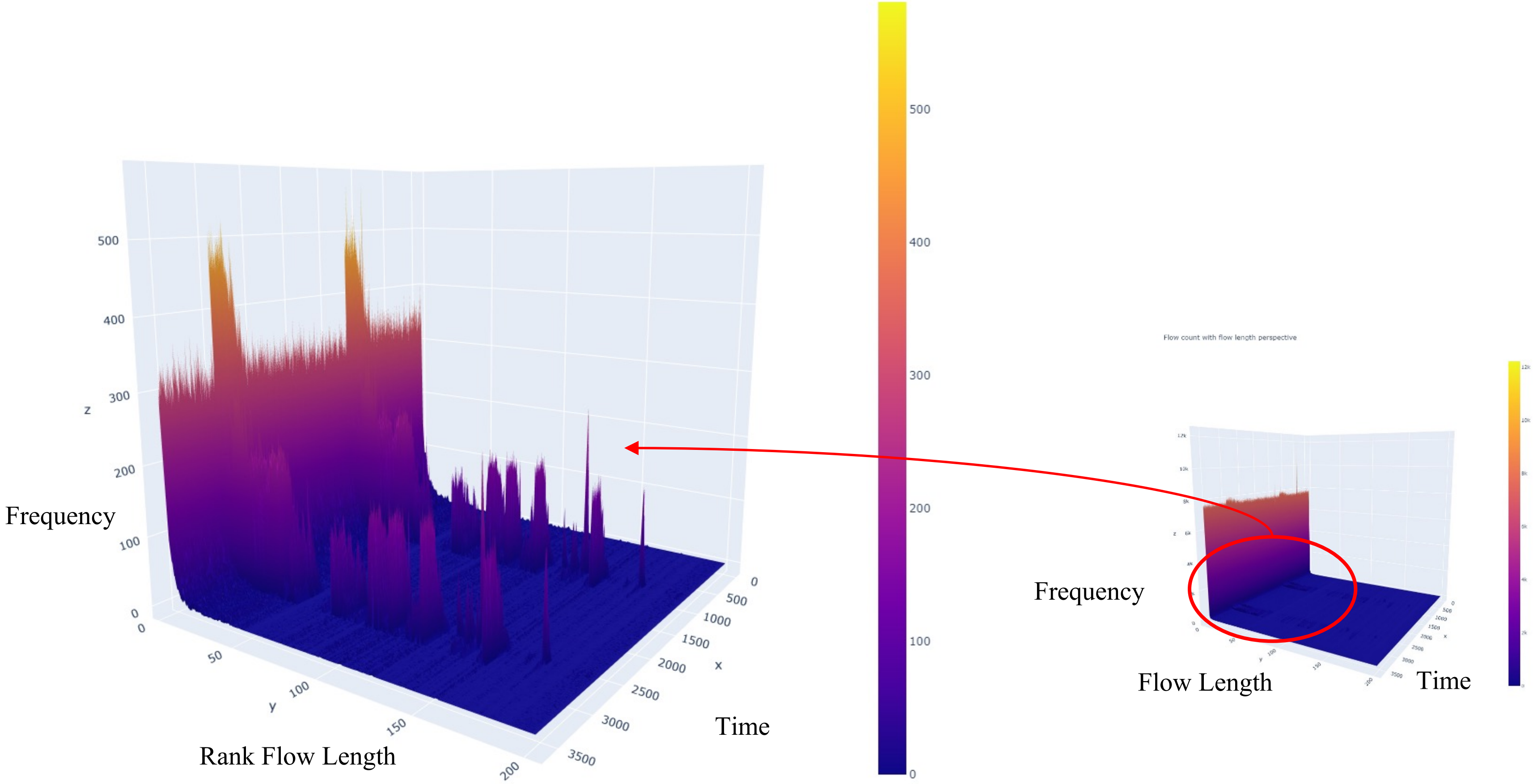
Scan Anomaly Detection

- **MAWI 20220110 traffic**
  - The fanout degree highlights the scan anomalies (upper panel)
  - The normalized second frequency moments of the source and destination IP addresses (lower panel)



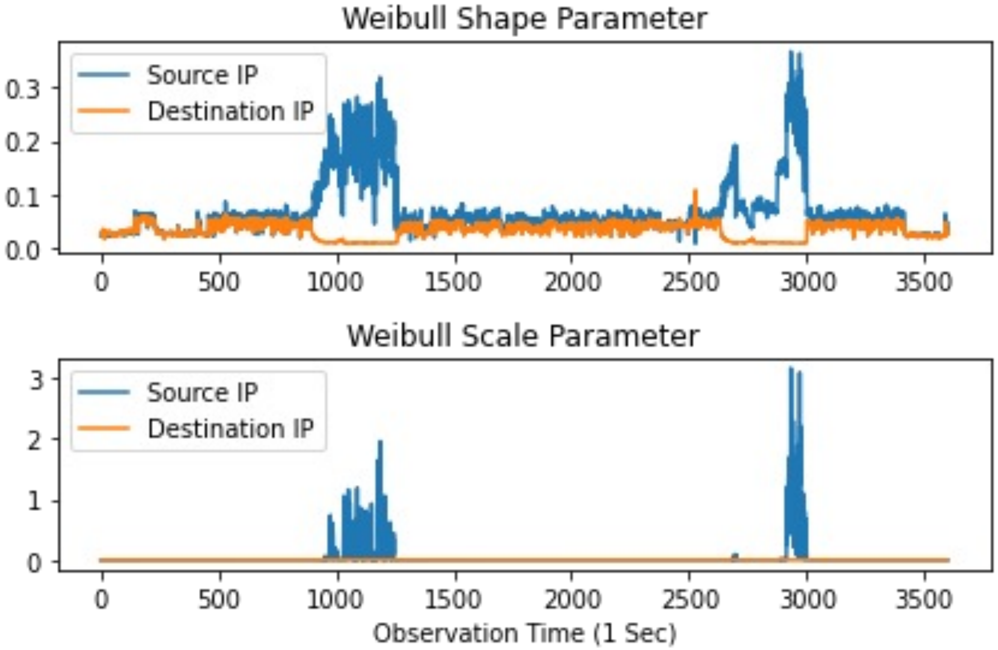
Flow Length Distribution

- Synthetic MAWAI + CAIDA 2007 DDoS traffic.



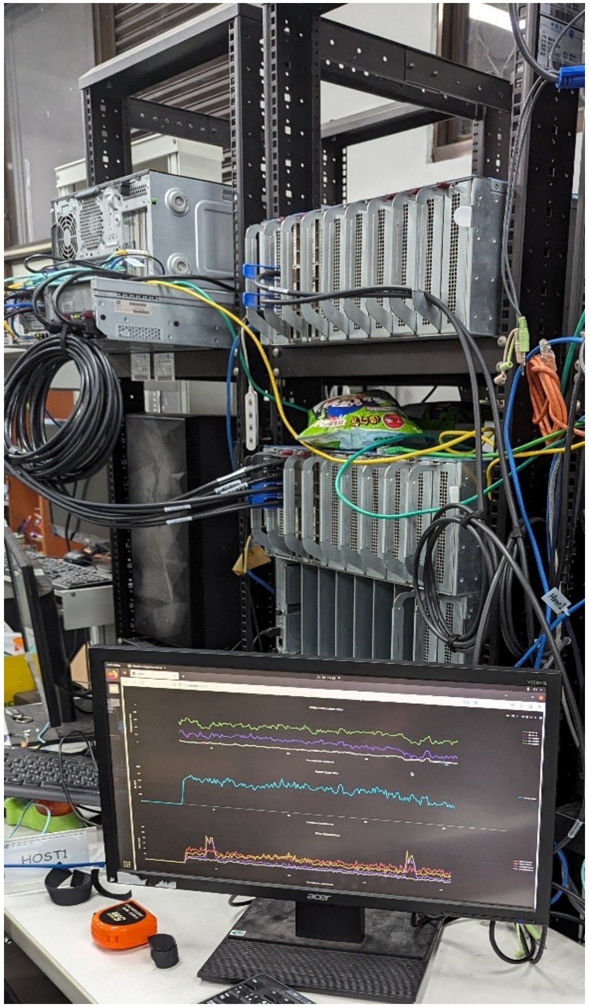
Weibull Model Parameter Estimation

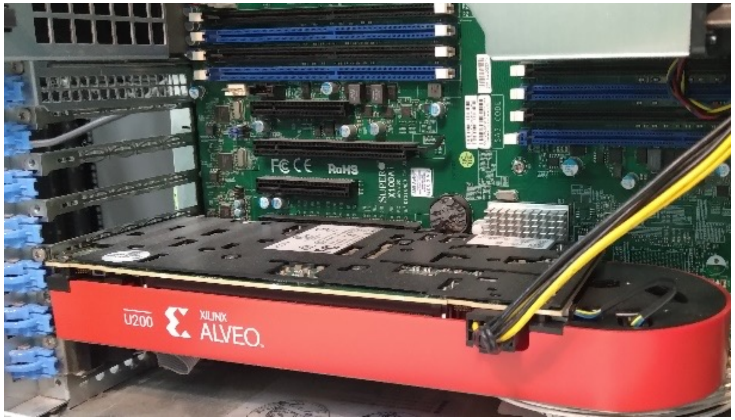
- **Method of Moment** **[11]**
- **Online estimation of Weibull parameters**
  - shape (k) and scale (λ)



[11] Ivana Pobocikova and Zuzana Sedliackova. 2014. Comparison of four methods for estimating the Weibull distribution parameters. *Applied Mathematical Sciences* 8 (2014), 4137–4149. https://doi.org/10.12988/ams.2014.45389

Testbed





Xilinx FPGA Demo

- Replays the synthetic trace from a 2-port 100Gbps NIC
- Observation time of 30 seconds.

Conclusion

- ❑ **Sketch-based Frequency Moment Estimation**
  - ○ Xilinx Alveo U200 FPGA
  - ○ NetFPGA PLUS Framework
- ❑ **Real-time online processing of network traffic**
- ❑ **2nd Frequency Moment Estimation**
  - ○ Scan Anomaly
  - ○ Weibull Parameter Estimation on Flow-Length distribution
- ❑ **Demo**