中原大學
Chung Yuan Christian University

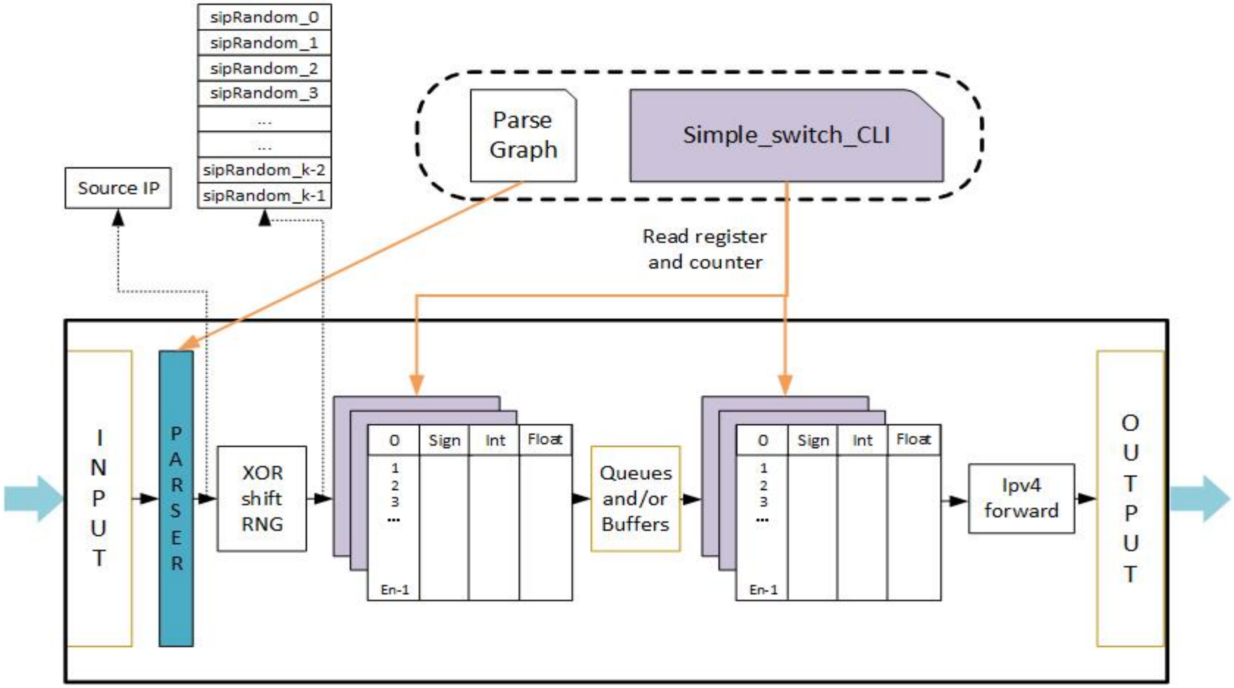
1st P4 European
Workshop (P4EU)
September 24, 2018

A Sketch-based Network Traffic Entropy Estimation using P4

Ku-Yeh Shih, Yu-Kuen Lai, He-Ping Li, Bo-Yu Huang and Yu-Jau Lin
Computer Network and Systems Research Laboratory, Department of Electrical Engineering
Chung Yuan Christian University, Zhongli 32023, Taiwan

- **Entropy usage**
  - Massive data mining, Machine learning
  - Traffic anomaly. DDoS, port scan[1]
- **Empirical entropy in real-time**
  - Limited memory and process time[2]
- **Estimation techniques to speed up**
  - Clifford and I. Cosma. Method[5], streaming
    - $R(i_t) = \tan(W_1)[\frac{\pi}{2} - W_1] + \log\left(W_2 \frac{\cos W_1}{\pi/_2 - W_1}\right)$
    - Transform into table lookup / Wire speed
- Total memory space
  - Use $k = 20$ tables of 64K($E_n = 2^{16}$) entries
  - 20*456KB $\cong$ 8.91MB
- **Accuracy tradeoff**
  - Quality of random number generator[3], sketch and table size

Trace name	Small	Medium	Large
# of packets	6,948,502	16,531,395	17,486,529
# of distinct SIP	79,823	98,933	183,933



Trace name	Small	Medium	Large
Relative error(%)	6.89%	0.37%	4.36%

[1] A. Lakhina, M. Crovella, and C. Diot, “Mining anomalies using traffic feature distributions,” SIGCOMM Comput. Commun. Rev., vol. 35, no. 4, pp. 217–228, 2005.

[2] A. Lall, V. Sekar, M. Ogihara, J. Xu, and H. Zhang, “Data streaming algorithms for estimating entropy of network traffic,” SIGMETRICS Perform. Eval. Rev., vol. 34, no. 1, pp. 145–156, 2006.

[3] G. Marsaglia, “Xorshift RNGs,” Journal of Statistical Software, vol. 8, no. 14, 2003.

[4] V. M. Zolotarev, One-dimensional Stable Distributions. American Mathematical Society, 1986.

[5] P. Clifford and I. Cosma, “A simple sketching algorithm for entropy estimation over streaming data,” in Artificial Intelligence and Statistics, 2013, pp. 196–206.