Persistent Spread Measurement for Big Network Data Based on Register Intersection*

You Zhou

Department of CISE
University of Florida, USA
youzhou@cise.ufl.edu

Min Chen

University of Florida
Google Inc., USA
min@cise.ufl.edu

Yian Zhou

University of Florida
Google Inc., USA
yian@cise.ufl.edu

Shigang Chen

Department of CISE
University of Florida, USA
sgchen@cise.ufl.edu

ABSTRACT

Persistent spread measurement is to count the number of distinct elements that persist in each network flow for pre-defined time periods. It has many practical applications, including detecting long-term stealthy network activities in the background of normal-user activities, such as stealthy DDoS attack, stealthy network scan, or faked network trend, which cannot be detected by traditional flow cardinality measurement. With big network data, one challenge is to measure the persistent spreads of a massive number of flows without incurring too much memory overhead as such measurement may be performed at the line speed by network processors with fast but small on-chip memory. We propose a highly compact Virtual Intersection HyperLogLog (VI-HLL) architecture for this purpose. It achieves far better memory efficiency than the best prior work of V-Bitmap, and in the meantime drastically extends the measurement range. Theoretical analysis and extensive experiments demonstrate that VI-HLL provides good measurement accuracy even in very tight memory space of less than 1 bit per flow.

CCS CONCEPTS

• **Networks** → **Network measurement**; **Network monitoring**; *Denial-of-service attacks*; *Network management*;

KEYWORDS

Persistent Spread Measurement; Big Network Data; Network Traffic Measurement; Network Security

*The full version of the paper is available at <https://www.cise.ufl.edu/~sgchen/paper/sigmetrics17.pdf>

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

SIGMETRICS '17, Urbana-Champaign, IL, USA

© 2017 Copyright held by the owner/author(s). ACM ISBN 978-1-4503-5032-7/17/06.

DOI: <http://dx.doi.org/10.1145/3078505.3078593>

ACM Reference format:

You Zhou, Yian Zhou, Min Chen, and Shigang Chen. 2017. Persistent Spread Measurement for Big Network Data Based on Register Intersection. In *Proceedings of SIGMETRICS '17, Urbana-Champaign, IL, USA, June 5-9, 2017*, 1 pages.

DOI: <http://dx.doi.org/10.1145/3078505.3078593>

Massive and distributed data are increasingly prevalent in modern networks as high-speed routers forward packets at hundreds of gigabits or even terabits per second. Big data also happens at the network edge. Traffic measurement and classification at such high speeds and with such massive volumes pose significant challenges.

Flow cardinality estimation is a fundamental problem in network traffic measurement. It estimates the number of *distinct* elements in every flow during pre-defined measurement periods. Each *flow* is uniquely identified by one or multiple fields in the packet headers, called *flow label*, which can be flexibly defined based on application needs. Existing research on flow cardinality estimation mainly focuses on analysing traffic sketches from one measurement period, which is the summary of the raw traffic data in that time period. This paper studies an under-investigated problem of analyzing sketches across multiple periods. In particular, we are interested in measuring the *persistent spread* of each flow, which is defined as the number of distinct elements that show up in a network flow during a certain number of consecutive measurement periods.

Our Contributions: First, we design a highly efficient persistent spread estimator called Intersection HLL (I-HLL) that works over multiple measurement periods. Second, to further improve memory efficiency, we introduce register sharing on top of I-HLL and propose a highly compact Virtual Intersection HLL (VI-HLL) architecture to measure the persistent spreads of a large number of flows simultaneously. Finally, the experimental results demonstrate the superior performance of VI-HLL.

ACKNOWLEDGMENTS

This work is supported in part by the National Science Foundation under grant STC-1562485 and a grant from Florida Center for Cybersecurity.