

Accurate Sampling of the Internet for Effective Anomaly Detection

<http://www.ece.ucdavis.edu/rubinet/sand.html>



Lead-PI: Chen-Nee Chuah, University of California, Davis

NSF CNS-0716831

Co-PI: Jun Xu, Georgia Institute of Technology

NSF CNS-0716423

Project Goals

- Characterize (negative) impact of sampling on anomaly detection
- Develop advanced sampling techniques to obtain accurate traffic statistics

Overview

Packet/flow sampling is used to collect traffic measurements in backbone network to cope with high data rate. Sampled data is sufficient for traditional management tasks (e.g., capacity planning and load balancing), but we show that it *reduces effectiveness of existing portscan & volume detection schemes*.

Our Approach

- Identify traffic features critical for signature- and non-signature based detection & quantify how much they are distorted by various sampling process
- Develop *Fast-Filtered Sampling (FFS)*, which combines low-complexity filter with existing sampling schemes to capture both large flows ('elephants') and small flows ('mice') such as portscan traffic that typically precedes virus propagation.

Approach and Impact

New approach

- Quantify how different packet/flow sampling impact portscan and volume anomaly detection
- Two-stage filter/sampler to control flow-size dependent sampling

Research Impact

- Identify 'flow-thinning' and 'flow-shortening' effect of sampling
- Accurately measure both 'elephants' & 'mice'
- Ensure effective anomaly detection

Technical Discussion

- Fast Filtered Sampling (FFS) is a light-weight implementation that achieves the same objectives as Sketch-Guided Sampling (SGS). FFS consists of 'Filter' module has an array of N counters of m bits each. Incoming packet is hashed based on flow-id to one of the counters. If counter value is $\leq s$, the packet is passed onto sampler, otherwise it is discarded. If counter value $\geq l$, it is reset to zero.

'Sampler' module performs random or uniform packet sampling with probability p .

- Sampling curve as a function of flow size:

$$\Pr \left\{ \begin{array}{l} \text{A packet sampled} \\ \text{from flow of size } i \end{array} \right\} = \begin{cases} p & \text{if } 1 \leq i < s \\ ps/i & \text{if } s \leq i < l \\ ps/l & \text{otherwise} \end{cases}$$

Selected Results

- Data collected via FFS, SGS, random packet sampling (RPS) and random flow sampling (RFS) is used for portscan detection using TAPS. All schemes achieve similar success ratio, but FFS & SGS achieve lower false positives than RPS.

