Uniformity and Independence of Hash Functions for Packet Sampling

Nick Duffield, AT&T
Saverio Niccolini, NEC
Maurizio Molina, DANTE
Juergen Quittek, NEC

Hash-based Packet Selection

- Router calculates hash of each packet
  - Input: $x \in$ hash domain = subset of invariant packet content
  - Output: $h(x) \in$ hash range R
- Router selects packet if $h(x) \in$ selection range $S \subset R$
- Required statistical properties
  - Assumption: need representative sample, e.g. estimate traffic rates
  - Want selection to resemble sampling at specified target rate p

    $\Rightarrow$ hash values h(x) should appear uniformly distributed: p = #S/#R
  - Want selection to resemble independent sampling between packets

    $\Rightarrow$ hash values $h(x_i)$ should appear independent over packets i
- This work
  - Evaluate hash uniformity and independence for PSAMP hash functions
    - Previous work examined uniformity, and execution speed.

Test Setup

- Hash Functions Evaluated
  - BOB, CRC32, IPSX, MMH
- Data
  - Packet traces from MAWI (wide area), and NEC (campus)
- Tests
  - Do packet hashes conform to uniform independent distribution?
  - Diverse Significance Tests
    - Chi-Square: uniformity
    - Box-Ljung: independence
    - Collision Test: uniformity and independence
- Dimensions
  - Dependence on hash input length
    - 16 bytes from IP/transport header + n bytes payload
  - Granularity of hash values (= minimum sampling rate considered)
  - Variability of results between traces

Summary of Results

- Lowest sampling rates considered is 1 in 1000
- Uniformity
  - BOB closest, CRC32 close to uniform, MMH and IPSX not close
  - BOB more consistent than CRC32 across different traces
  - BOB needs 20 bytes input (header + 4 bytes payload) in worst case
- Independence
  - Only BOB appears close to independent in all tests (20 byte input)
  - CRC32 matches BOB performance in some tests, but needs more input
- Execution speed
  - IPSX nearly order of mag. faster than BOB, which is faster than CRC32
    - IPSX potentially useful if scant resources in router, some applications