Packet Sampling (PSAMP)

Tanja Zseby, FhG FOKUS
zseby@fokus.fhg.de
Overview

- Initiative started by AT&T labs in autumn 2001
- IETF WG since summer 2002
- Goals
  - Description of sampling methods for
    - Configuration of sampling schemes and parameters
    - Reporting of sampling schemes and parameters to applications
  - Protocol for exporting (parts of) packets
- Close cooperation with IPFIX
Differences to IPFIX

- **IPFIX:**
  - Flow information export
  - Focus on export protocol
  - Configuration not considered
  - No flow selection

- **PSAMP:**
  - Packet export
  - Focus on configuration
  - Description of packet selection schemes
  - IPFIX chosen as basis for PSAMP protocol
PSAMP and IPFIX

Packet Capturing

Sampling

Filtering

Classification & Flow Recording

Flow Sampling

Flow Filtering

Observation Point:
- Router
- Probe

Exporting process

PSAMP

packet reports

IPIFX

flow records

flow records

both steps may be trivial (1:1 sampling, no filtering)
Current Drafts

- Framework draft-ietf-psamp-framework-10.txt
  - Overall Framework, Problem Statement
  - Example applications
  - Requirements for packet reporting
  - Status: IESG review

- Packet Selection draft-ietf-psamp-sample-tech-07.txt
  - Terminology
  - Packet selection methods
  - Sampling Parameters
  - Status: IESG review

- Information Model draft-ietf-psamp-info-02.txt
  - Status: Expired, waiting for IPFIX

- PSAMP MIB draft-ietf-psamp-mib-04.txt
  - Status: Waiting for info model

- PSAMP Protocol draft-ietf-psamp-protocol-01.txt
  - Status: Expired, waiting for IPFIX
Terminology

• Try to be consistent with IPFIX terminology
  – some common terms (e.g. observation point, flow, etc.)
  – use different terms if meaning is different e.g.
    • packet stream: (sub)set of packets (e.g. output of selection process)
    • flow: set of packets with common properties

• Filtering: deterministic selection based on the packet content

• Sampling: everything else
  – Content-independent Sampling
    • Deterministic or random selection independent of packet content
    • Examples: systematic, random sampling that is independent of packet content.
  – Content-dependent Sampling
    • Random selection dependent on packet content

• Mandatory to implement one of the described schemes in packet selection draft
Schemes and Parameters: Sampling

• Systematic count-based
  – start and stop in accordance to spatial packet position (packet count).
  – Input parameters:
    • Interval length (in number of packets)
    • Spacing between intervals (in number of packets)

• Systematic time-based
  – start and stop in accordance to temporal packet position (arrival time).
  – Input parameters:
    • Interval length (in µsec)
    • Spacing (in µsec)

• Not covered:
  – Systematic sampling with combined time- and count based trigger
  – Non-equal spacing
Schemes and Parameters: Sampling

• Random n-out-of-N
  – Random selection of n packets from N
  – Input parameters:
    • Population size N
    • Sample size

• Uniform Probabilistic
  – Same sampling probability for each packet
  – Input parameters:
    • Sampling probability p

• Non-Uniform Probabilistic
  – Sampling probability depends on input
  – Function for calculation probability p ➔ Vendor dependent

• Flow State Probabilistic
  – Sampling probability depends on flow state (of own flow or other flows)
  – Policy for selecting flows ➔ Vendor dependent
Schemes and Parameters: Filtering

- **Field Match**
  - Filtering according to fields described as IPFIX information elements
  - Concatenation of filters: only AND, no NOT or OR
  - Input Parameters
    - Fields
    - Selection range(s)

- **Hashing**
  - Apply bit mask to packet header and/or the first N bytes of the payload, create unique bit string
  - Optionally, link with another pre-defined bit string (seed)
  - Apply hash function on the string
  - Select the packet if the result falls into one or more a selection range(s)
  - Input Parameters
    - Input mask (as bit strings)
    - Seed
    - Selection interval
    - Hash-function
Schemes and Parameters

• Router-state selection
  – Select packet on the basis of its route/treatment in the router (e.g. IF to which it is routed, no route found, etc.)
  – Router state (when packet should be selected) ➔ up to vendor

• Combined Schemes
  – Combination of the defined sampling and filtering schemes
  – E.g. for stratified sampling
  – Coupled via STREAM_ID

[your scheme here]
<table>
<thead>
<tr>
<th>Selection Scheme</th>
<th>deterministic</th>
<th>content-</th>
<th>Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>selection</td>
<td></td>
<td>dependent</td>
<td></td>
</tr>
</tbody>
</table>

| systematic           | X             | _        | Sampling      |
| count-based          |               |          |               |

| systematic           | X             | -        | Sampling      |
| time-based           |               |          |               |

| random               | -             | -        | Sampling      |
| n-out-of-N           |               |          |               |

| random               | -             | (X)      | Sampling      |
| uniform probabilistic|               |          |               |

| random               | -             | (X)      | Sampling      |
| non-uniform probabil.|               |          |               |

| random               | -             | (X)      | Sampling      |
| non-uniform flow-state|              |          |               |

| field match          | X             | X        | Filter        |

| hash function        | X             | X        | Filter        |

| router state filter  | X             | (X)      | Filter        |
Hash-based Selection

• Hash-based Selection
  – Emulation of Random Sampling:
    • Synchronization of sampling processes at multiple observation points
  – Can be used for Packet Digest:
    • Packet identifier to correlate packet arrivals at different observation points

• Mandatory Hash-Function: BOB
  – Selected based on criteria
    • uniformity (for random emulation)
    • collisions (for digesting)

• IF hash-based sampling is implemented BOB MUST be used
  – But it is not mandatory to implement a hash-based selection to be PSAMP compliant

• Further Optional
  – CRC32
  – IPSX
Open Issues and Next Steps

- IPv6 input for hash functions
- Packet IDs
- PSAMP protocol
  - Requirements from framework document
  - Based on IPFIX
- Further schemes
Thank You