Packet Hashing and Sampling Applications

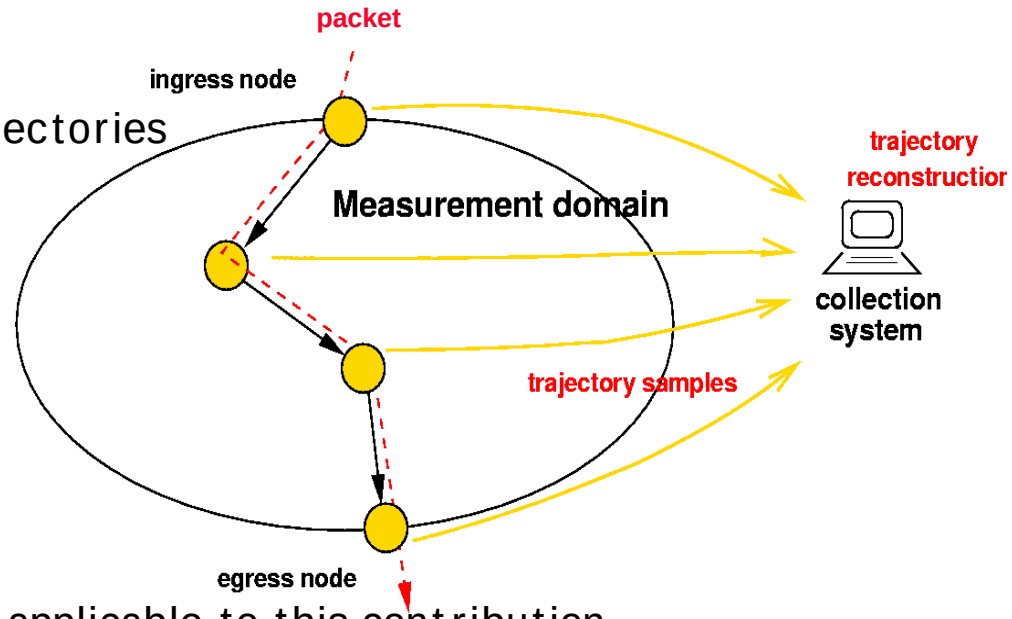
Nick Duffield

AT&T Labs

nduffield@att.com

Trajectory Sampling

- Basic Idea
  - router calculates hash (digest) of invariant packet fields
    - invariant = fields that don't change per hop; exclude TTL, checksum
  - select packet if hash falls in a specified range
  - each router uses same hashing function
- Consequence: Trajectory Sampling
  - each packet is sampled either **everywhere** or **nowhere**
- Reporting and Trajectory Reconstruction
  - routers report samples to collector
  - direct reconstruction of packet trajectories
    - from reported samples **alone**
  - no routing state info required
    - sidesteps network state uncertainty
      - latency/synchronization
      - randomization/load balancing
  - in real time



- N.B.
  - AT&T may own intellectual property applicable to this contribution

Applications

- Network engineering
  - map traffic flows onto network topology
  - actual traffic intensity = sampled traffic intensity / sampling rate
- Performance measurement
  - trajectory terminating in core ⇒ packet loss
- Real-time anomaly detection
  - self-intersecting trajectories ⇒ routing loop
- Network probing
  - specify packet content so that it is sampled

Standardization Issues

- Multi-vendor domains
  - packet selection depends on choice of hashing function
  - therefore need common hashing function across domain
- Need realistic requirements
- Hashing function requirements
  - more computational cycles ⇒ better hash function
    - hash appears uniformly distributed
    - good specification of sampling rate
  - implementation issues: needs computational resources
- Hashing function input requirements
  - more packet fields used as input ⇒ better statistical properties
    - hash appears uncorrelated with any given packet field
    - sampling decisions appear statistically random
  - implementation issues: needs fields available

Related packet hashing application

- IP traceback (Snoeren et. al.)
  - aim: trace path of packet with spoofed source IP address
  - each node: calculate multiple hashes of each packet
  - store compactly (Bloom filter)
  - upon network attack: collect filters centrally
  - attempt match suspicious packet against each filter
  - use matches to identify packet path
- Role for PSAMP
  - provider of packet measurements
  - router compute hashes
  - exports them locally to on-board IP traceback application