Distributed Iceberg Detection with SDN-enabled Online Learning

Chang Liu, Shu Ming Peng, Mehdi Malboubi, Chen-Nee Chuah, Matt Bishop, Ben Yoo
University of California at Davis

This work was partially supported by BBN under the GENI 4 subcontract 1953 under NSF CNS-1346688

➢ Network Measurement

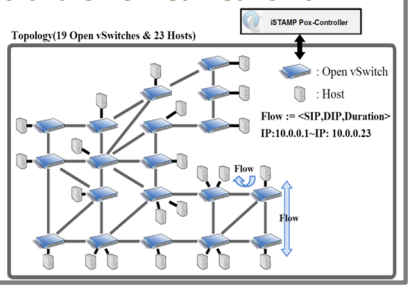
- Accurate and timely traffic matrix (TM) measurements provide essential inputs for today's various network operations, such as traffic engineering, capacity planning, network troubleshooting and anomaly detection.

➢ Software Defined Networking (SDN)

- SDN is considered as a promising element to implement traffic monitoring, management and control.
  - Separation of control plane from data plane
  - Centralized controller with a network-wide view enabling global optimization
  - Dynamically reprogram switches in a timely manner

➢ Network Tomography with Online Learning in Software Defined Networks

- Revisit the traffic matrix estimation (TME) problem in OpenFLow-based network.
  - Limited routing entries & measuring entries
- Utilize the online learning feasibility provided by SDN to target measuring "important" flows
- Heuristic solutions to allocate "important" flows to distributed TCAM resources



➢ Network Tomography with Online Learning in Software Defined Networks

- Objective:
  - Our goal is to estimate the network-wide traffic matrix by collaboratively managing the distributed measurement resources (TCAM entries) in the network.
- Solution:
  - Install per-flow measurements in the measuring entries to avoid aggregation and routing feasibility issues
  - Use an intelligent online learning algorithm to sample the most important flows
  - Utilize the online learning feasibility provided by SDN to update the measuring entries periodically to target the most important flows
  - Heuristic solutions to allocate "important" flows to distributed TCAM resources

Algorithm 1 Lasthop

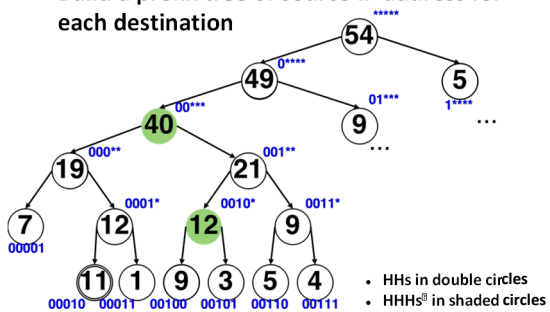
Input: index=[ranking flows based on their "importance"], distribution of measuring entries, assume k_j TCAM entries available at switch j
Output:a feasible switch-flow allocation
for each flow fl in index **do**
  switch_list = [switches flow fl goes through from dst to src]
  for each switch j in switch_list **do**
    if $k_j > 0$ **then**
      install per-flow measurement for flow fl here
      $k_j = k_j - 1$
      break
    end if
  end for
end for

Algorithm 1 Greedy

Input: index=[ranking flows based on their "importance"]; distribution of measuring entries, assume k_j TCAM entries available at switch j; load for each switch: number of important flows passes trough this switch
Output: a feasible switch-flow allocation
for each flow fl in index **do**
  switch_list = [switches flow fl goes through from dst to src]
  choose switch $j \in switch\_list$ where k_j is largest
  if there is a tie **then**
    choose switch j which has the least load
  end if
  if $k_j > 0$ **then**
    install per-flow measurement for flow fl here
    $k_j = k_j - 1$
    break
  end if
end for

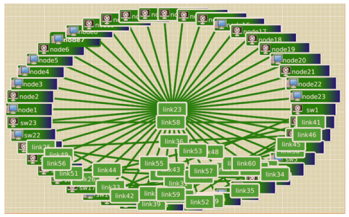
➢ Simulation Results

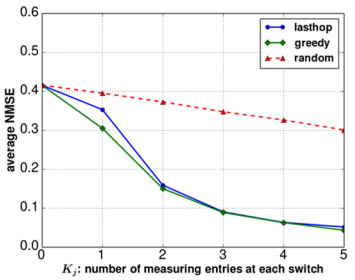
- Simulate our framework using GEANT topology
  - 23 switches, 37 links
  - Real traffic traces of GEANT network is used
- Application of TME
  - Heavy Hitter (HH) Detection
    - Flows with a flow size larger than a threshold
  - Hierarchical Heavy Hitters (HHH) Detection
    - Build a prefix tree of source IP address for each destination



- HHs in double circles
- HHHs[B] in shaded circles

[B]Jose, Lavanya, Minlan Yu, and Jennifer Rexford. "Online measurement of large traffic aggregates on commodity switches." *Conference on Hot Topics in Management of Internet, Cloud, and Enterprise Networks and Services–Hot-ICE. USENIX.* 2011.

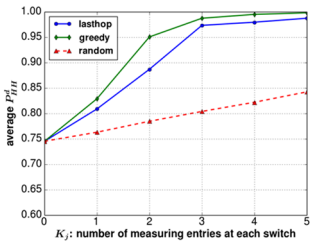


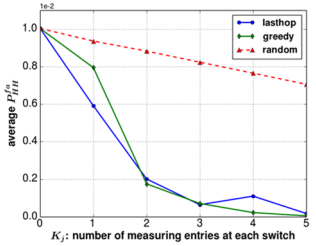


- K_j: number of measuring TCAM entries at each switch
- NMSE measures the accuracy of traffic matrix estimation

➢ Global Iceberg Detection

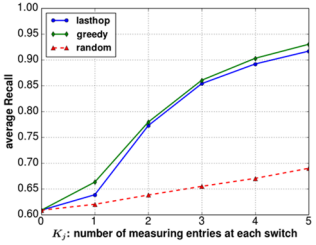
- Global Heavy Hitter (HH) Detection

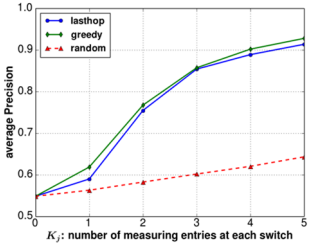




- Threshold: 10% of link capacity
- P^d_{HH}: probability of detection
- P^{fa}_{HH}: probability of false alarm

- Global Hierarchical Heavy Hitters (HHH) Detection





- Threshold: 10% of link capacity
- Recall: the total number of true HHHs detected over the real number of HHHs
- Precision: the total number of true HHHs detected over the total number of HHHs reported