INTERNATIONAL CONFERENCE ON PARALLEL PROCESSING

ICPP/2021/CHICAGO/USA

acm In-Cooperation

sighpc

AUGUST 9-12, 2021

sRouting: Towards a Better Flow Size Estimation Performance through Routing and Sketch Configuration

Yang Shi (时洋) and Mei Wen

National University of Defense Technology

August 12, 2021

Outline

- Motivation
- sRouting Overview
- Offline Part
- Online Part
- Evaluation

Outline

- **Motivation**
- sRouting Overview
- Offline Part
- Online Part
- Evaluation

Motivation

国防科技大学
National university of defense technology

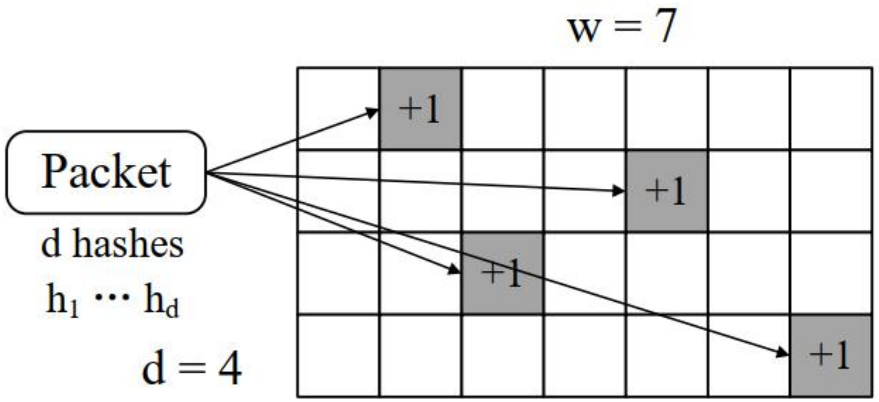
- Sketches are widely used for flow size estimation
  - Emerging programmable switches
  - Small amount of memory
  - Fine-grained statistics

- Adopting sketches in real-life facing practical issues
  - Partial deployment of programmable devices
  - Unique characteristics and objectives in different networks

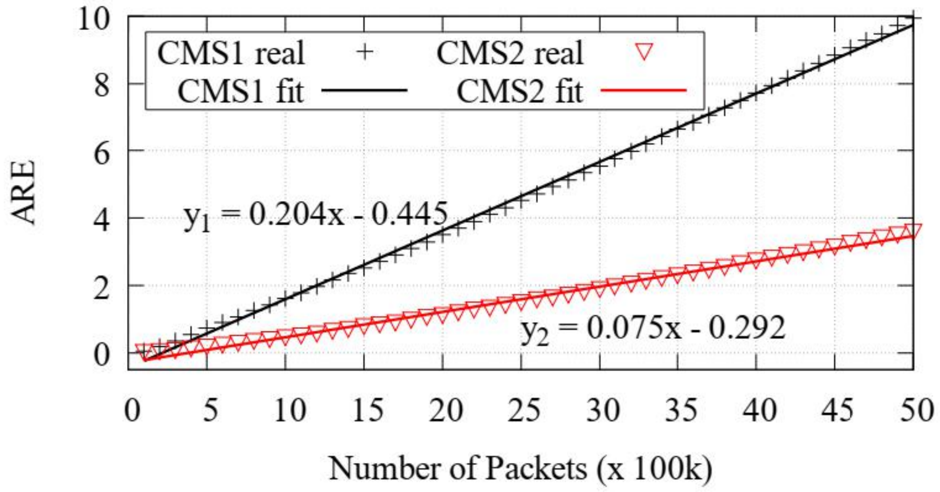
How to improve the performance of sketches ?

Analysis of sketches

- Accuracy of sketches
  - Using ARE (Average Relative Error)
  - In theory： $\hat{f_i} \leq f_i + \epsilon \|f\|_1$
  - Testing results of different sketches
    - Can be approximated well with linear functions



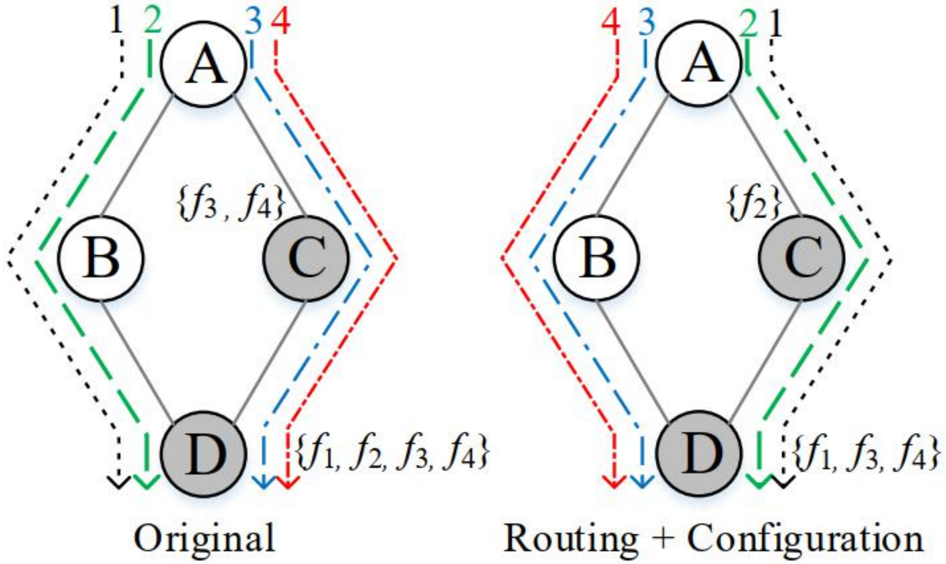
Count-Min sketch example



Count-Min sketch accuracy

Example

- Accuracy sketch-based measurement can be improved
  - using routing and sketch configuration



Sketch measurement in networks

Routing table in switch C

Match	Action	
Flow Key	Next Hop	Sketch
f_1	D	0
f_2	D	1

Routing table in switch D

Match	Action	
Flow Key	Next Hop	Sketch
f_1	-	1
f_2	-	0
f_3	-	1
f_4	-	1

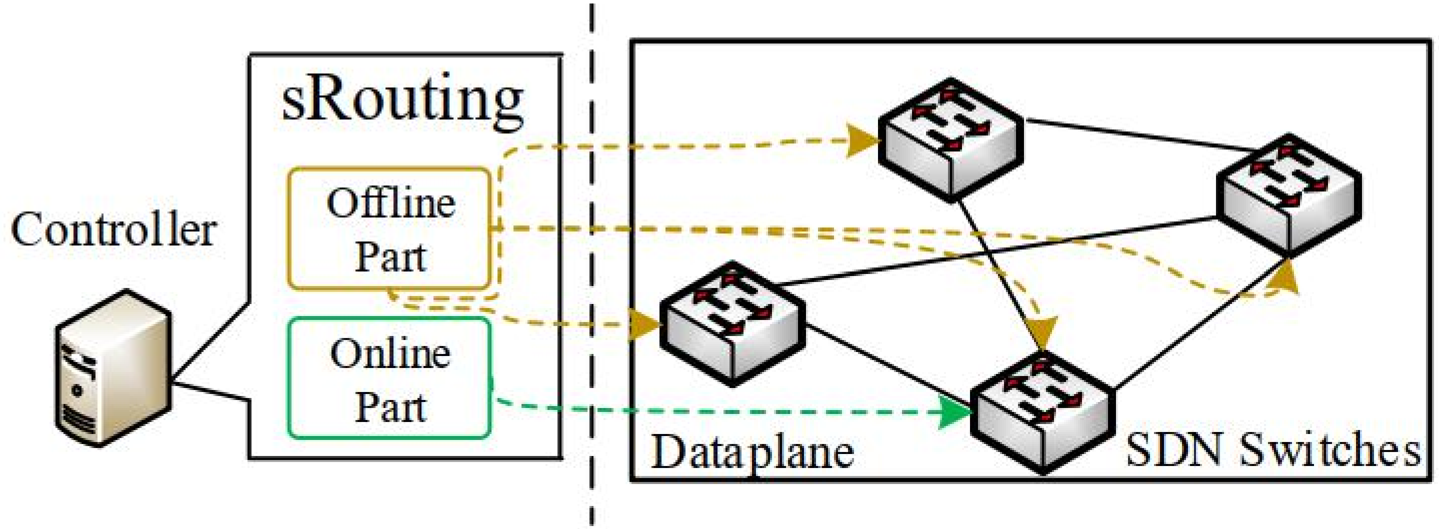
Routing table in switch C and D

Outline

- Motivation
- **sRouting Overview**
- Offline Part
- Online Part
- Evaluation

sRouting

- Two parts:
  - Offline part, making strategies on the base of long-term information
  - Online part, adjusting strategies when facing dynamic changes



Outline

- Motivation
- sRouting Overview
- **Offline Part**
- Online Part
- Evaluation

Network and sketch models

- Network
  - Flow, flowset, macroflow
  - Flows in a flowset shares one routing path

- Sketch
  - Partial deployment of programmable switches (sketches)
  - Different measurement importance values

- Makes sRouting Practical and general

Offline problem formulation

- ILP model for offline scenario

$$(\mathrm{O})\ \min \sum_{\gamma_i \in \Gamma} u(\gamma_i) q_{\gamma_i} \quad (2)$$

$$s.t. \sum_{p \in P_{\gamma_i}} x_{\gamma_i}^p = 1, \qquad \forall \gamma_i \in \Gamma \quad (3)$$

$$\sum_{v_k \in p} y_{\gamma_i}^k = x_{\gamma_i}^p, \qquad \forall \gamma_i \in \Gamma \quad (4)$$

$$\sum_{\gamma_i \in \Gamma} \sum_{e \in p : p \in P_{\gamma_i}} f_{\gamma_i} \le c(e), \qquad \forall e \in E \quad (5)$$

$$\sum_{\gamma_i \in \Gamma} \sum_{p \in P_{\gamma_i}} x_{\gamma_i}^p I_p^k \le c(t), \qquad \forall v_k \in V \quad (6)$$

$$N_{v_k} = \sum_{\gamma_i \in \Gamma} y_{\gamma_i}^k N_{\gamma_i}, \qquad \forall v_k \in V \quad (7)$$

$$q_{v_k} = (1 - Z_k)T + Z_k(\alpha_s N_{v_k} + \beta_s), \qquad \forall v_k \in V \quad (8)$$

$$q_{\gamma_i} = \sum_{v_k} y_{\gamma_i}^k q_{v_k}, \qquad \forall \gamma_i \in \Gamma \quad (9)$$

Randomized Rounding Algorithm

- Two steps:
  - Solving a relaxed LP model
  - Obtain integer solutions
- With proven performance
  - Both constraints and measurement performance

Algorithm 1: Randomized Rounding Algorithm for Offline sRouting

Input: ILP formulation with integer variables $x^p_{\gamma_i}, y^k_{\gamma_i} \in \{0, 1\}$

Output: Solutions $\{\hat{x}^p_{\gamma_i}, \hat{y}^k_{\gamma_i}\}$ of the original problem

Step 1: Solving the relaxed LP problem

Obtain the solutions $\widetilde{x}^p_{\gamma_i}, \widetilde{y}^k_{\gamma_i}$ of the optimization problem by relaxing all integer variables

Step 2: Deciding the routing path and measurement strategies

for *flowset* γ_i **do**
    Generate a value $\lambda_1 \in (0, 1)$
    for $p \in P_{\gamma_i}$ **do**
        $\lambda_1 = \lambda_1 - \widetilde{x}^p_{\gamma_i}$
        if $\lambda_1 < 0$ **then**
            $\hat{x}^p_{\gamma_i} = 1$
            Generate a value $\lambda_2 \in (0, 1)$
            for $v_k \in p$ **do**
                $\lambda_2 = \lambda_2 - \widetilde{y}^k_{\gamma_i}$
                if $\lambda_2 < 0$ **then**
                    $\hat{y}^k_{\gamma_i} = 1$
                    break
            break

Return $\{\hat{x}^p_{\gamma_i}, \hat{y}^k_{\gamma_i}\}$

Outline

- Motivation
- sRouting Overview
- Offline Part
- **Online Part**
- Evaluation

Online part of sRouting

- Main challenge in online scenario
  - Deployed strategies already in the network

- Formulated as a multi-objective problem
  - O1, minimal modification in deployed strategies
  - O2, maximize the network throughput
  - O3, minimize the measurement error

$$\textbf{(O1)}\ \min \sum_{\gamma_i \in \Gamma} \sum_{\psi \in \Psi_{\gamma_i}} z_{\gamma_i}^{\psi} g(\psi'_{\gamma_i}, \psi)$$

$$\textbf{(O2)}\ \max \sum_{\gamma_i \in \Gamma} \sum_{\psi \in \Psi_{\gamma_i}} z_{\gamma_i}^{\psi} f_{\gamma_i}$$

$$\textbf{(O3)}\ \min \sum_{\gamma_i \in \Gamma} u(\gamma_i) q_{\gamma_i}$$

Greedy algorithm for online sRouting

国防科技大学
National university of defense technology

- Detailed in Algorithm 2
  - Step 1, finding the flowsets which need adjustment
  - Step 2, finding the new scheme for important flowsets
  - Step 3, trying to accept more important floesets through replacing others

Algorithm 2: Online Algorithm in sRouting

Input: Γ, C_{max}, Ψ_{γ_i}, $u(\gamma_i)$, ψ'_{γ_i}
Output: ψ_{γ_i} for $\gamma_i \in \Gamma$

//Step 1: Finding the flowsets which need adjustment
Sort Γ in a decreasing order of $u(\gamma_i)$
for γ_i *in sorted* Γ **do**
  if ψ'_{γ_i} *is still available* **then**
    $\psi_{\gamma_i} = \psi'_{\gamma_i}$
  else
    $\Gamma_r \leftarrow \Gamma_r \cup \{\gamma_i\}$
    $\psi_{\gamma_i} =$
    $C_{max} = C_{max} - g(\psi'_{\gamma_i}, \psi_{\gamma_i})$
//Step 2: Finding new scheme for $\gamma_i \in \Gamma_r$
for γ_i *in* Γ_r **do**
  Sort Ψ_{γ_i} in an increasing order of q_ψ
  for $\psi_m \in \Psi_{\gamma_i}$ **do**
    if ψ_m *is available and* $C_{max} \geq g(\psi'_{\gamma_i}, \psi_m)$ **then**
      $\psi_{\gamma_i} = \psi_m$
      $\Gamma_r \leftarrow \Gamma_r \setminus \{\gamma_i\}$
      $C_{max} = C_{max} - g(\psi'_{\gamma_i}, \psi_{\gamma_i})$
      break
//Step 3: Trying to accept more important flowsets
for γ_i *in* Γ_r **do**
  $\Gamma_i = \{\gamma | \psi_\gamma \neq, u(\gamma) < u(\gamma_i)$, γ and γ_i belong to the same macroflow$\}$
  for $\psi_m \in \Psi_{\gamma_i}$ **do**
    for $\gamma_j \in \Gamma_i$ **do**
      $\psi_{\gamma_j} =$
      if ψ_m *is available for* ψ_{γ_i} *and* $C_{max} \geq g(\psi'_{\gamma_i}, \psi_m) + g(\psi'_{\gamma_j}, \psi_{\gamma_j})$ **then**
        $\psi_{\gamma_i} = \psi_m$
        $\Gamma_r \leftarrow \Gamma_r \setminus \{\gamma_i\}$
        $\Gamma_r \leftarrow \Gamma_r \cup \{\gamma_j\}$
        $C_{max} = C_{max} - g(\psi'_{\gamma_i}, \psi_{\gamma_i}) - g(\psi'_{\gamma_j}, \psi_{\gamma_j})$
        break
      $\psi_{\gamma_j} = \psi'_{\gamma_j}$
    if $\psi_{\gamma_i} \neq$ **then**
      break
Return $\{\psi_{\gamma_i}\}$

Outline

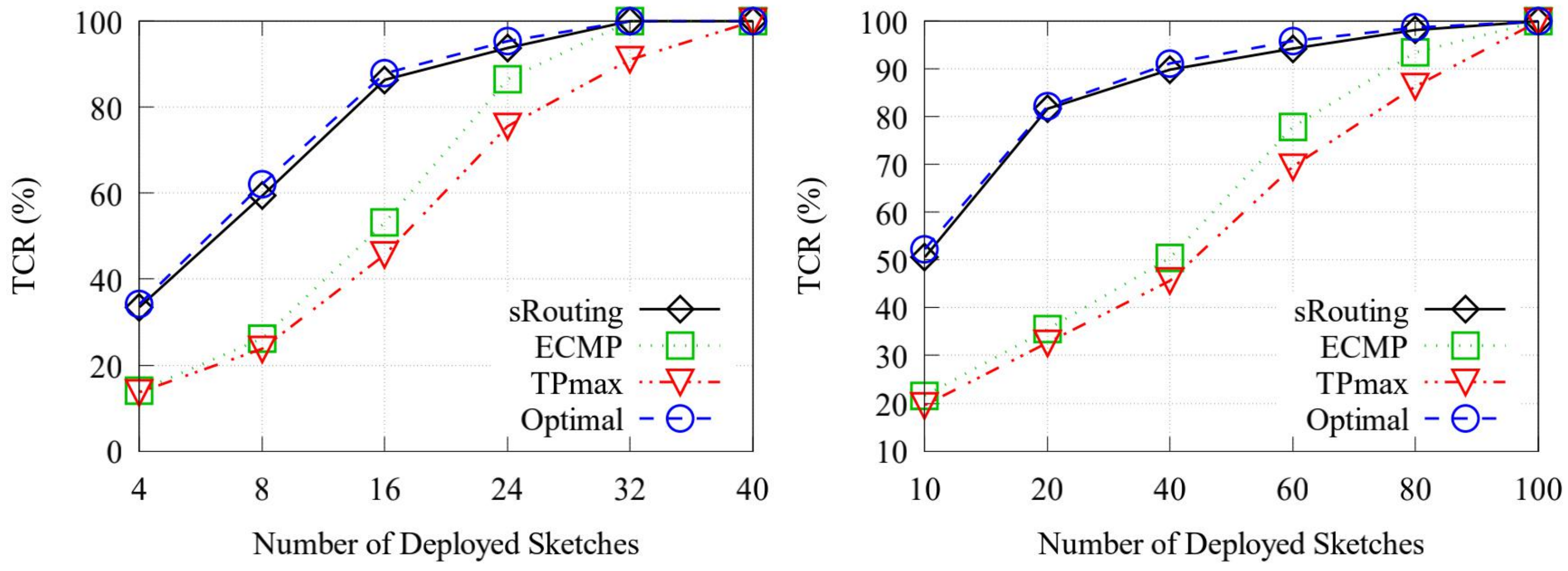
- Motivation
- sRouting Overview
- Offline Part
- Online Part
- **Evaluation**

Evaluation

- Simulation settings
  - Topology: FatTree with 80 switches and Monash with 50 switches
  - Sketch: Count-Min sketch, 600 KB
  - Trace: CAIDA network trace
  - Flowset: each macroflow is divided into 10 flowsets
- Metrics and baselines
  - Metrics: ARE, TCR (traffic cover ratio) and throughput
  - Baselines: Optimal, ECMP and TPmax

Offline scenario

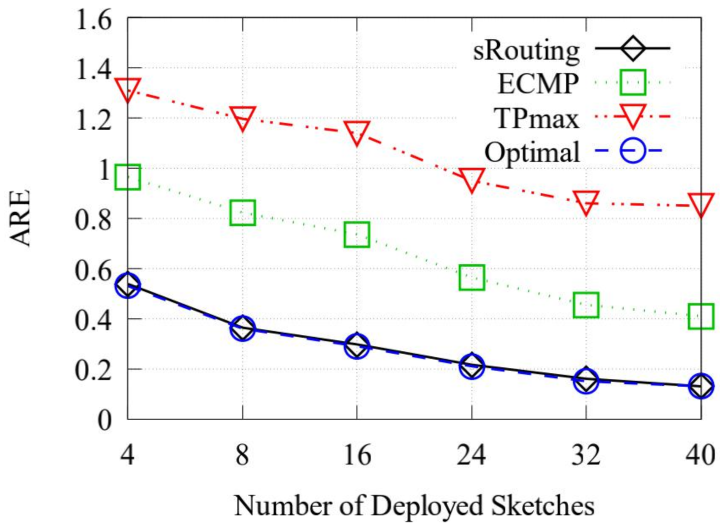
- TCR
  - FatTree (sketch = 8,16), 1.62 - 2.27x more traffic than ECMP and TPmax
  - close to Optimal
  - schedule significantly more traffic across sketches



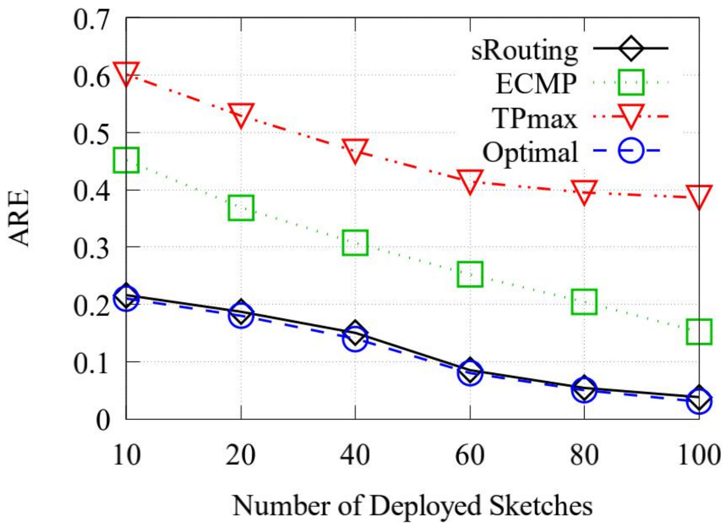
(a) TCR of FatTree network (b) TCR of Monash network

Offline scenario

- ARE
  - 40.5% and 26.3% that of ECMP and TPmax respectively
  - sRouting can achieve a much lower ARE
  - use as many sketches as possible to record the flows



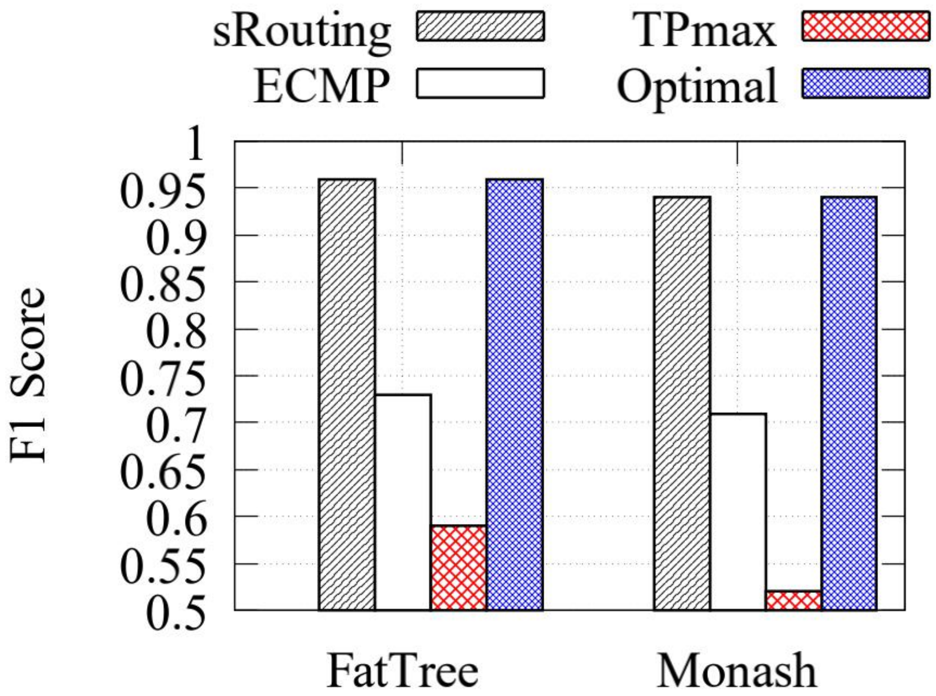
(c) ARE of FatTree network



(d) ARE of Monash network

Offline scenario

- heavy hitter detection
  - a higher value of F1 represents more accurate heavy hitter detection
  - F1 scores are improved by ratios of 31.5% and 62.7% (FatTree)
  - verify the effectiveness in improving the performance of other tasks



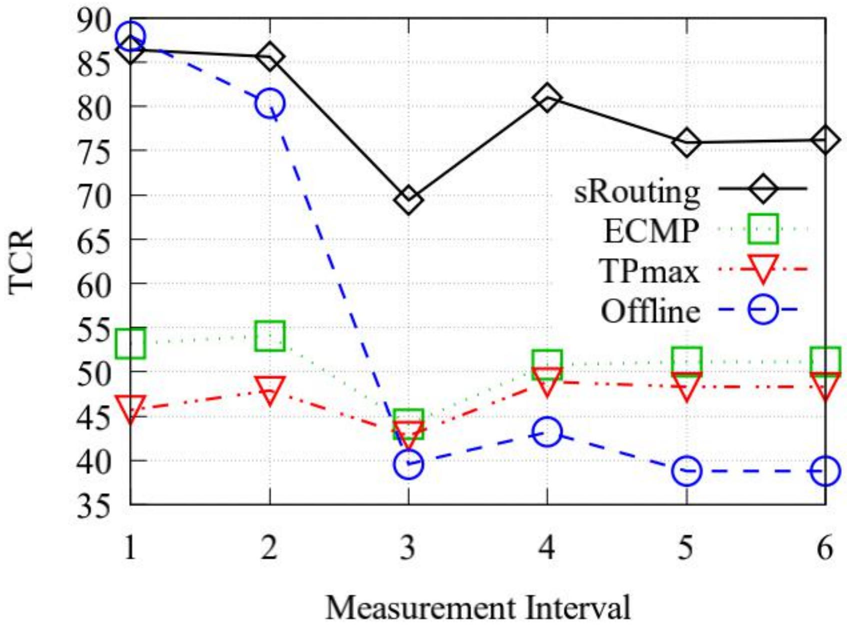
Online scenario

- Dynamic events in online evaluation

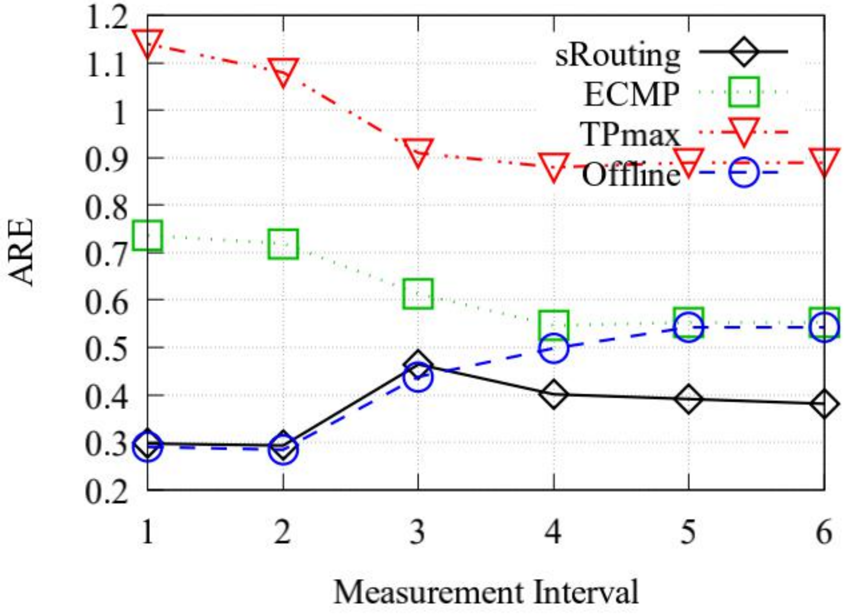
Event	Details
E_1	Remove five links from the topology
E_2	Double the traffic rates of all flows
E_3	Deploy five more sketches in the network
E_4	Exchange the traffic rate of 50% of elephant flows with the same number of mice flows
E_5	Change the importance value of 5% of elephant flows with the same number of mice flows

Online evaluation

- TCR and ARE
  - sRouting has a great effect on keeping the TCR at a high level
  - a small increase in the ARE of sRouting, much lower than others



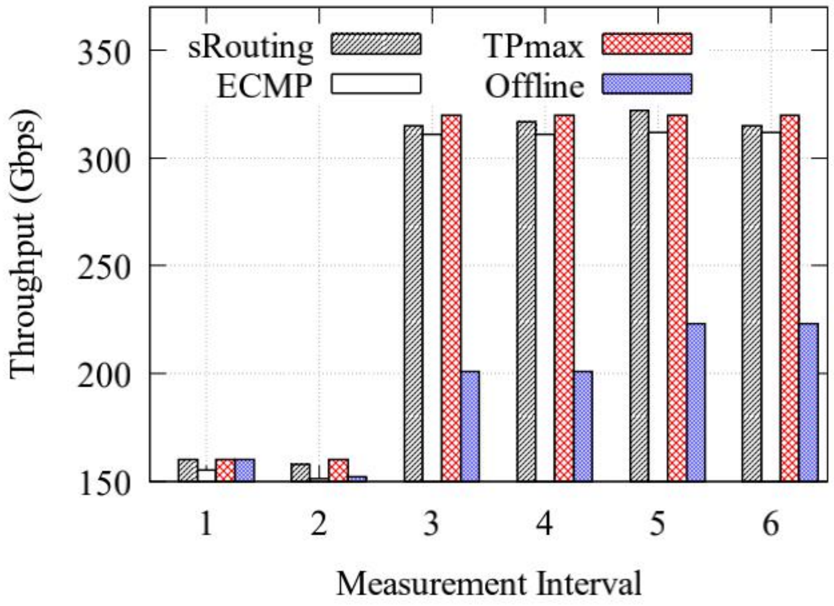
(a) TCR of FatTree network



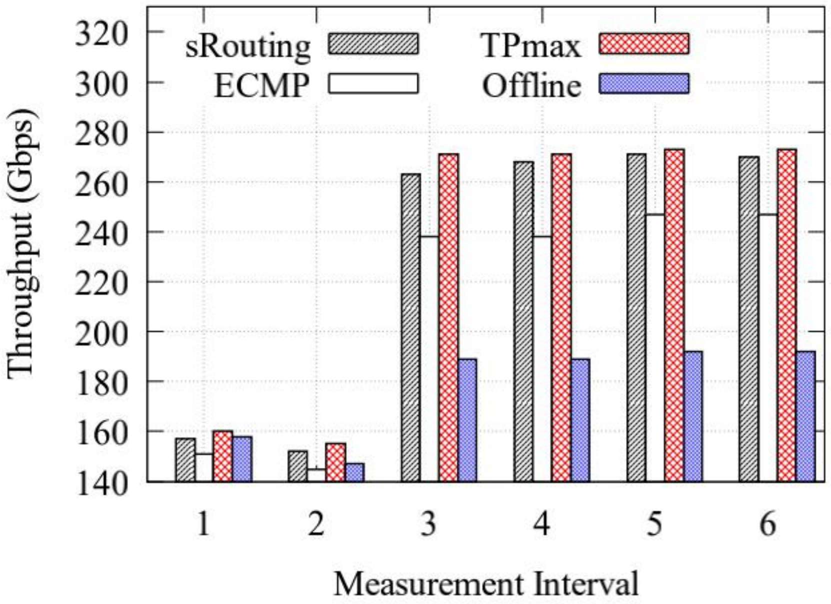
(b) ARE of FatTree network

Online scenario

- Throughput
  - TPmax achieves the highest throughput in most intervals
  - sRouting can achieve similar throughput to TPmax
  - with a maximum difference of about 2.9%



(c) Throuoghput of FatTree network



(f) Throuoghput of Monash network

Thank you !

谢谢!