iCAN (instant Congestion Assessment Network) for Traffic Engineering

(draft-liu-ican-01)

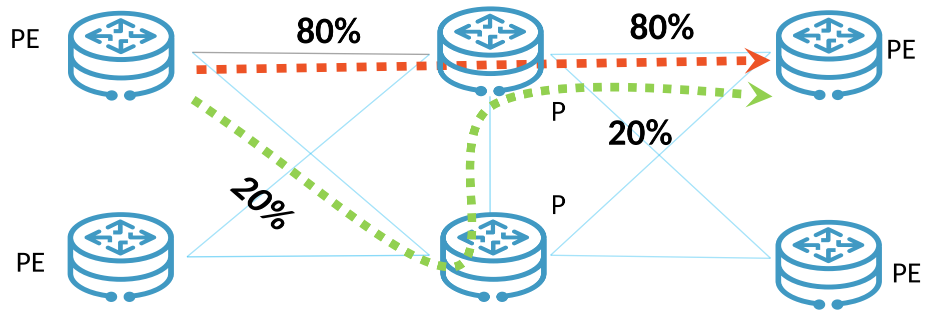
Joanna Dang(Huawei), liubing(Huawei),
Guangming Yang(China Telecom), Kyungtae Lee(LG U+)
@teas, ietf106

Contents

- Problem: Current Network is Unbalanced
  - Coarse Granularity of Traffic Adjustment
  - Slow Response (Ignorant to Microburst Traffic)
- Proposed Solution: iCAN Technologies
  - Flow Granularity of Traffic Adjustment
  - Millisecond-level Closed Control Loop
- Discussion

Problem: Current Network is Unbalanced

Partial Network Congestion



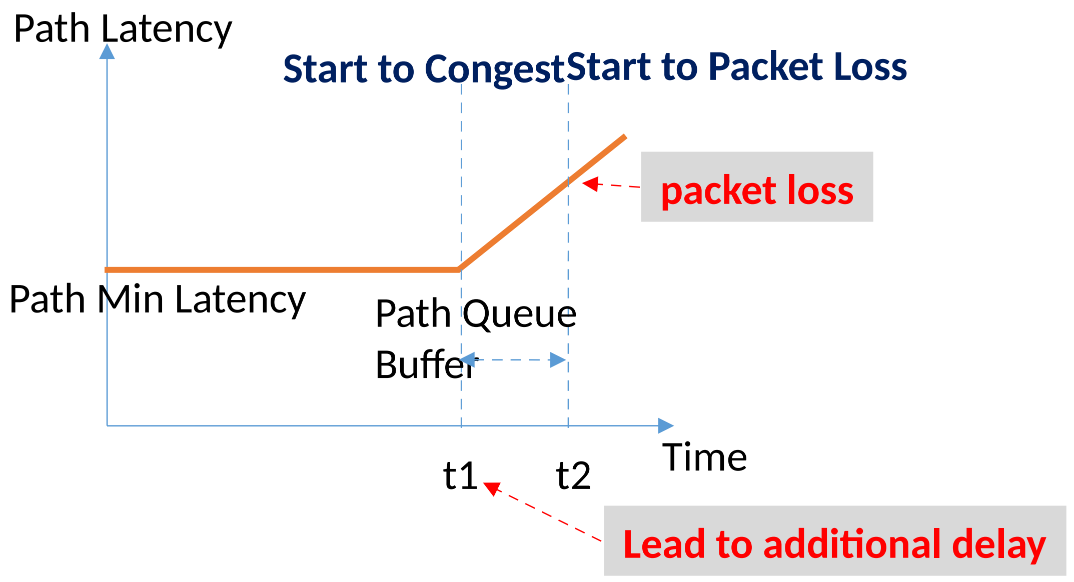
Low network throughput

- The average bandwidth utilization of the most operators' network is approximately 30%.

Bad user experience

- Unnecessary congestion leads to more packet loss and more delays

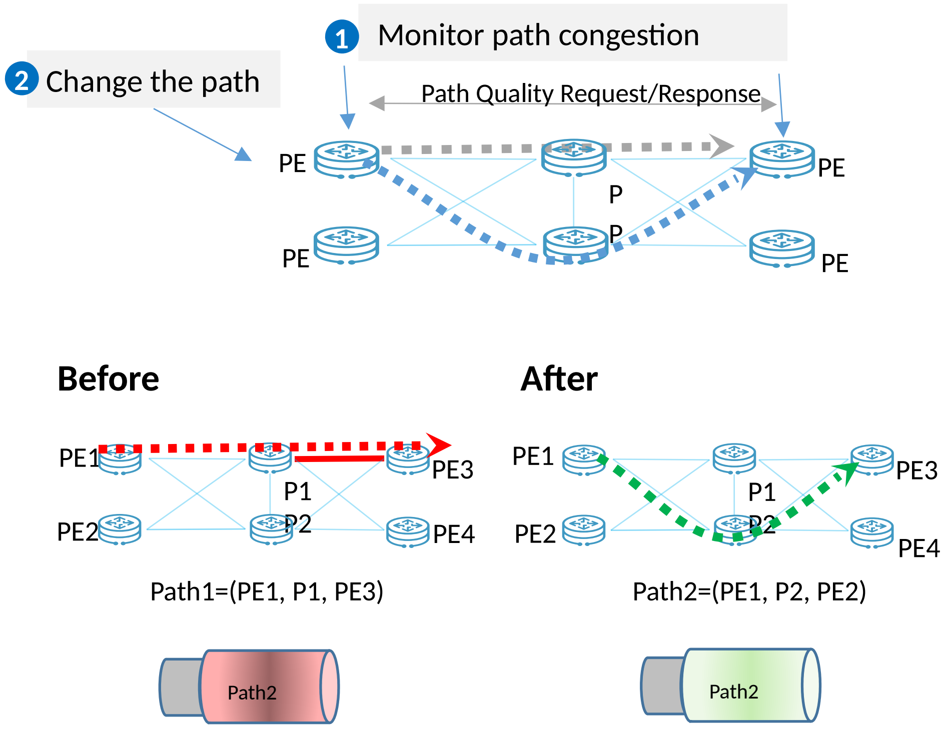
Packet loss and increased delay



- In general, the device buffer could tolerant some of the network congestion, avoiding packet loss caused by path congestion. But this mechanism leads to additional delay which could seriously impact low latency services.
- When device buffer is exceeded, packet loss would still occur.

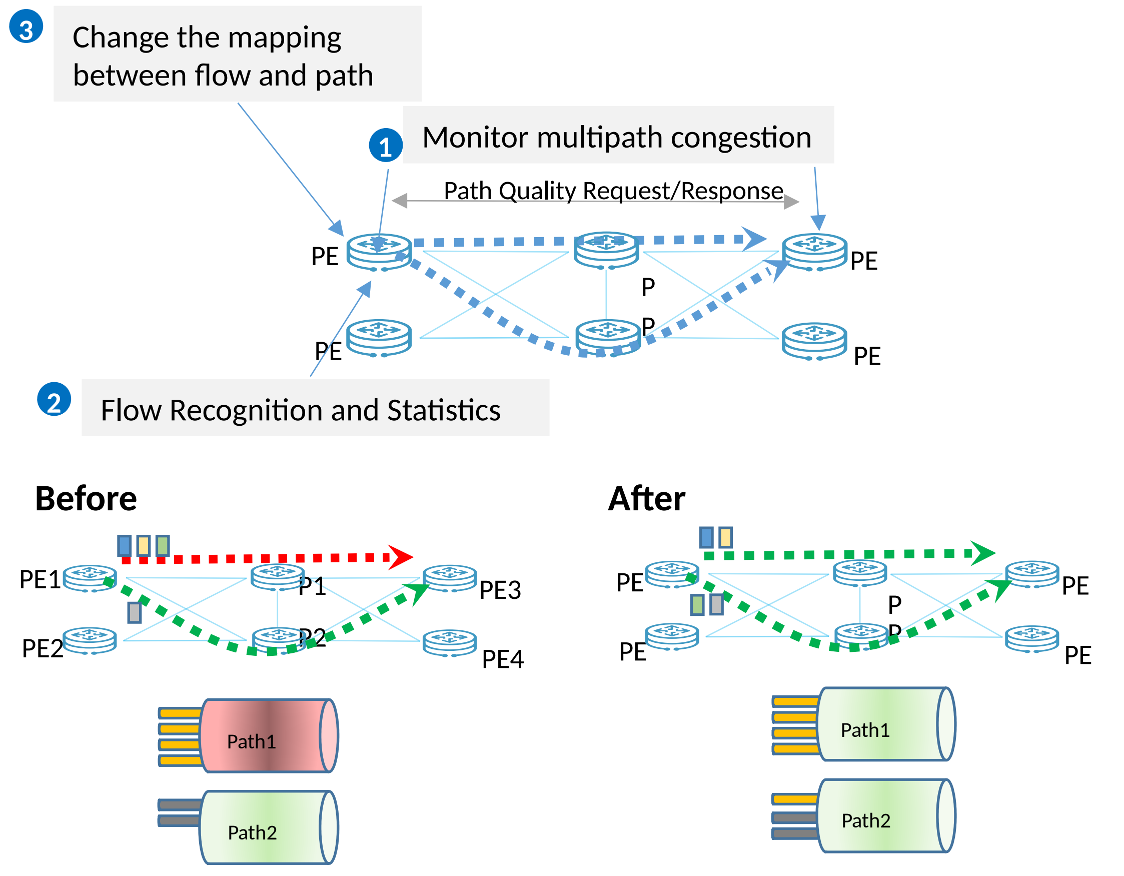
Why (1/2) : Coarse Granularity of Traffic Adjustment

Path Level



The adjustment object is path. All the flows carried by the path are faced with path switching.

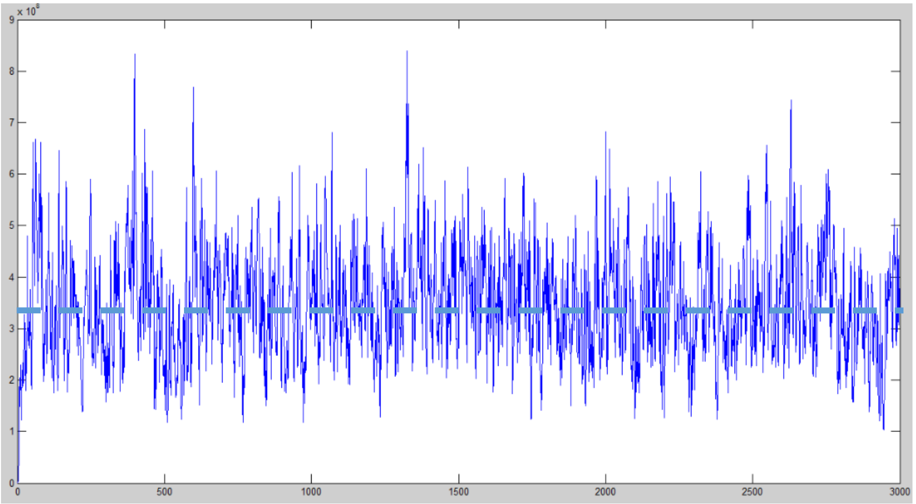
New: + Flow Level



The adjustment object is flow. Only some flows carried by the path are faced with path switching.

Why (2/2) : Microburst Issue

Path Traffic

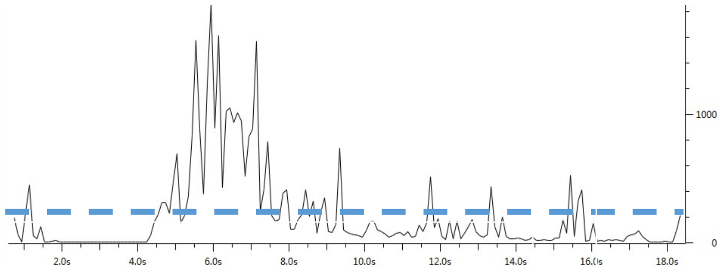


3000 seconds

Peak-to-Average ratio: >=3~6

(Aggregated traffic)

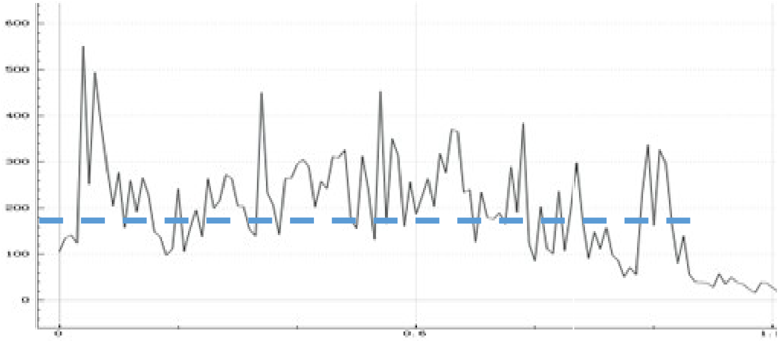
Example 1: Interactive Business (Such as Amazon and Taobao)



Peak-to-Average ratio: 86

1.9Gbps/22Mbps

Example 2: Video Business (Such as iQiyi and Vibrato)

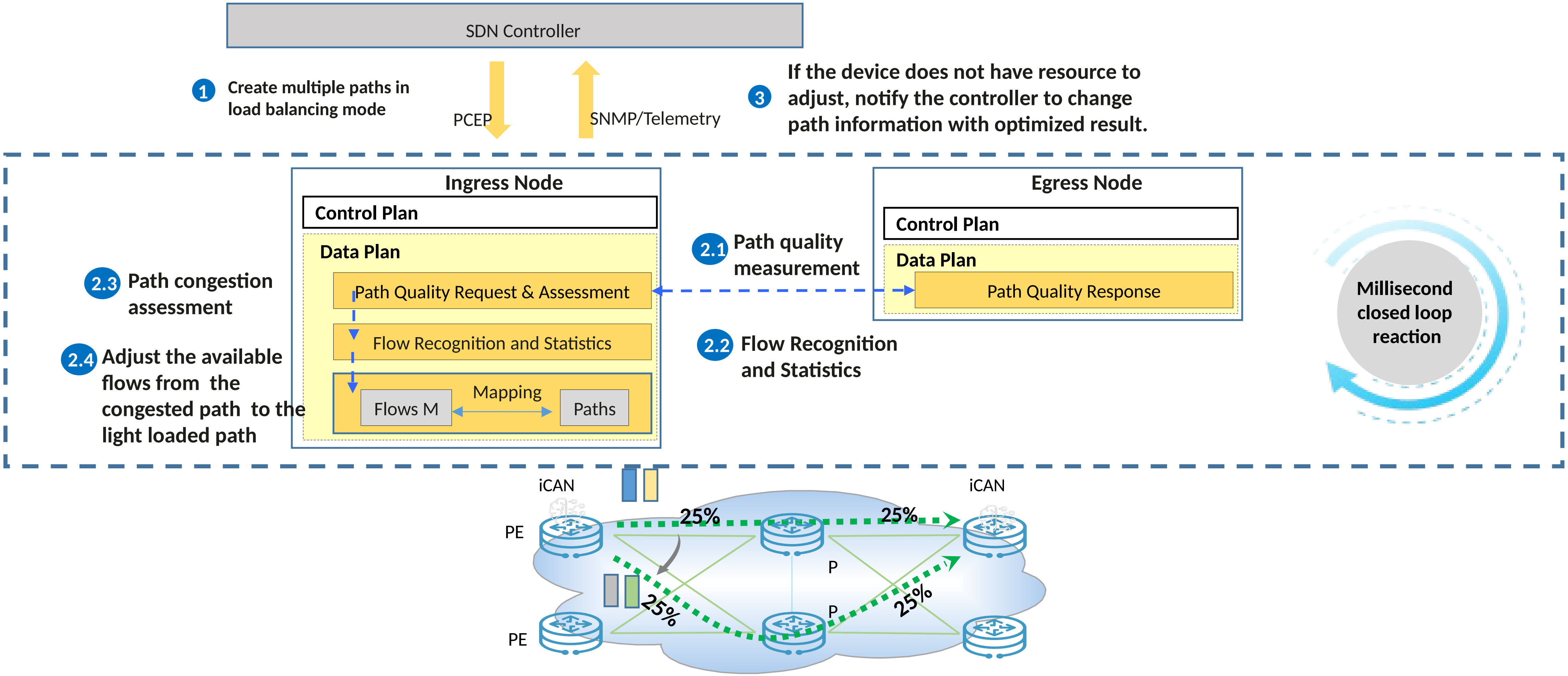


Peak-to-Average ratio: 5.7

528Mbps/93Mbps

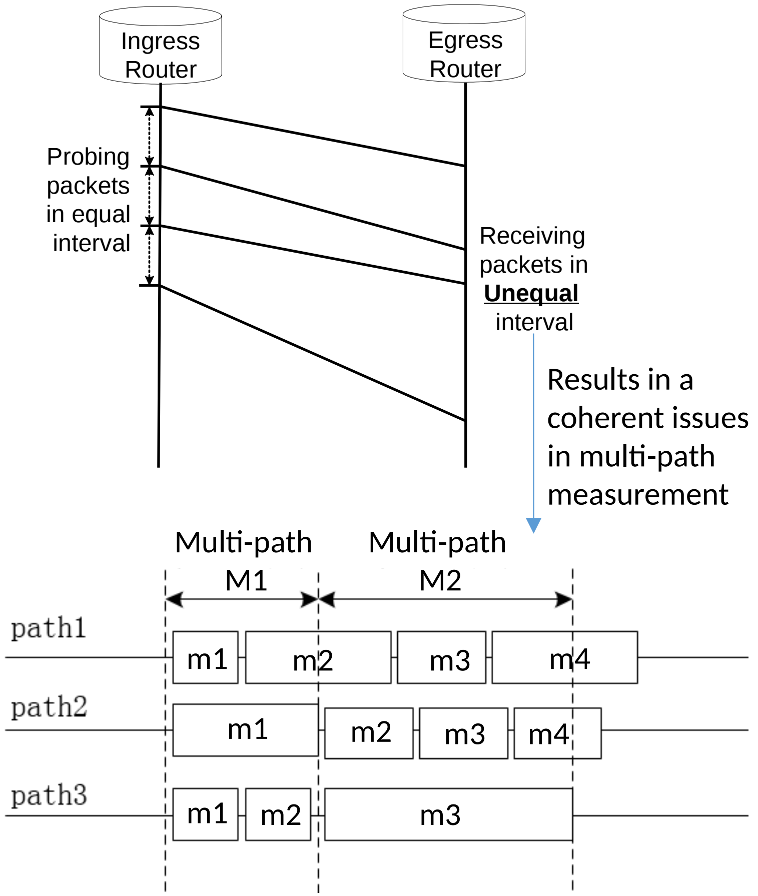
With the continuous introduction of new interactive services, the peak-to-average ratio of network traffic will become more and more serious.

Proposed Solution: iCAN (instant Congestion Assessment Network)



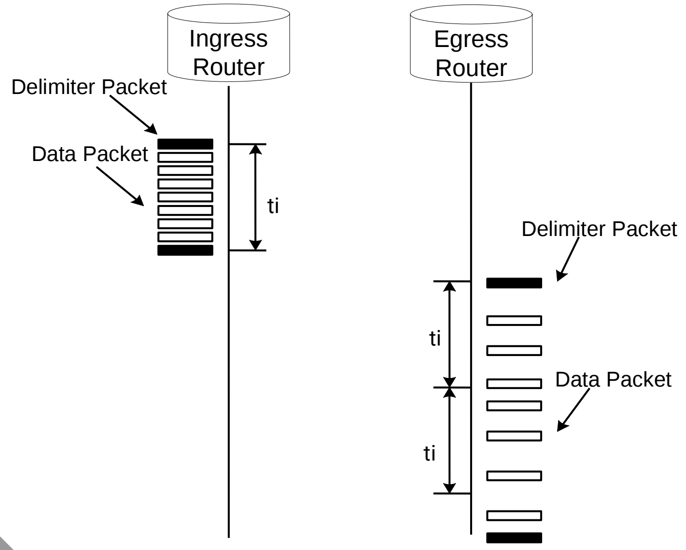
Key technologies: Path quality assessment - *Coherent multi-path measurement*

Problem

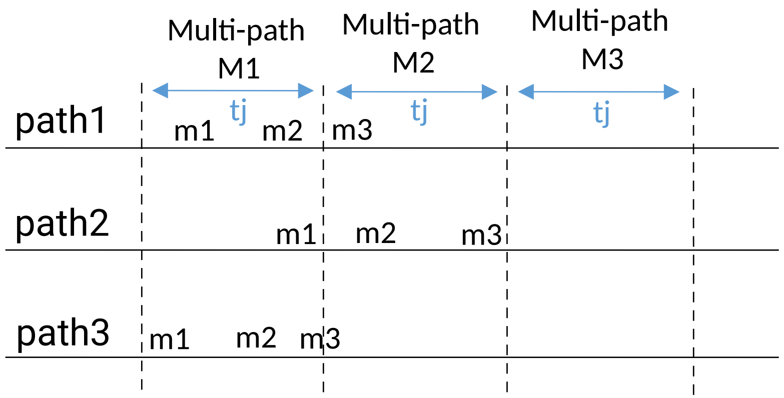


- On each path, the egress router would feedback the measurement results (m1, m2…) according to its own real interval.
- The ingress router would have to wait until the last m1/m2/m3 of the latest path come back.

Solution

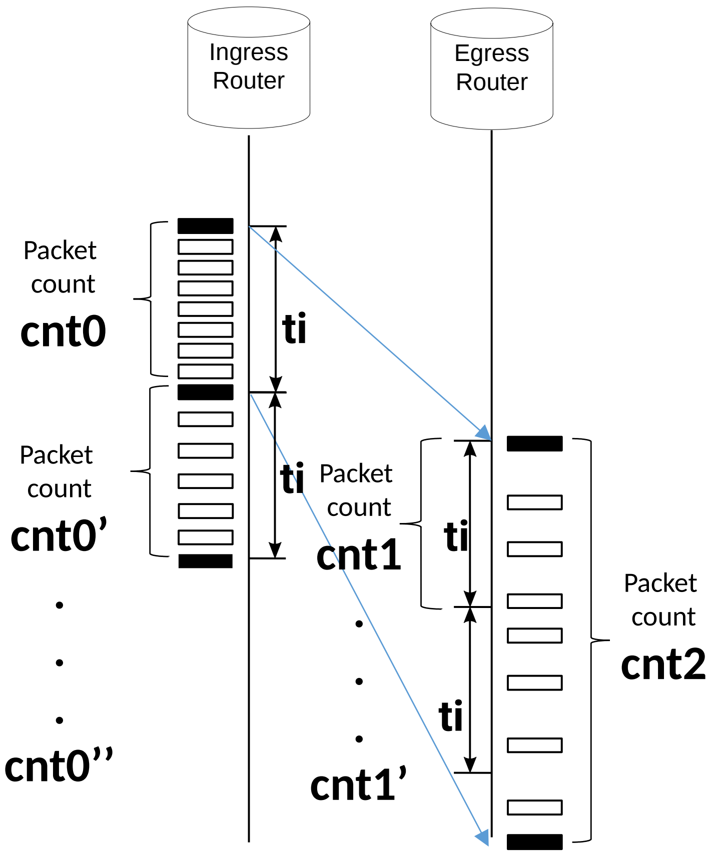


- The active probing packet acts as the delimiter packet among normal data packets. (In current prototype, the probing packet would be sent every 3.3ms, e.g. ti=3.3ms)
- Regardless of the shifting of the probing packets, the egress router would return the measurement result to the ingress router every ti internal.



- The ingress router would assess each path's congestion status every tj interval (In current prototype, tj=10ms)
- Tj should be larger enough than ti, so that every tj interval, the ingress would get at least one measurement result of each path.

Key technologies: Path quality assessment - *path congestion evaluation*



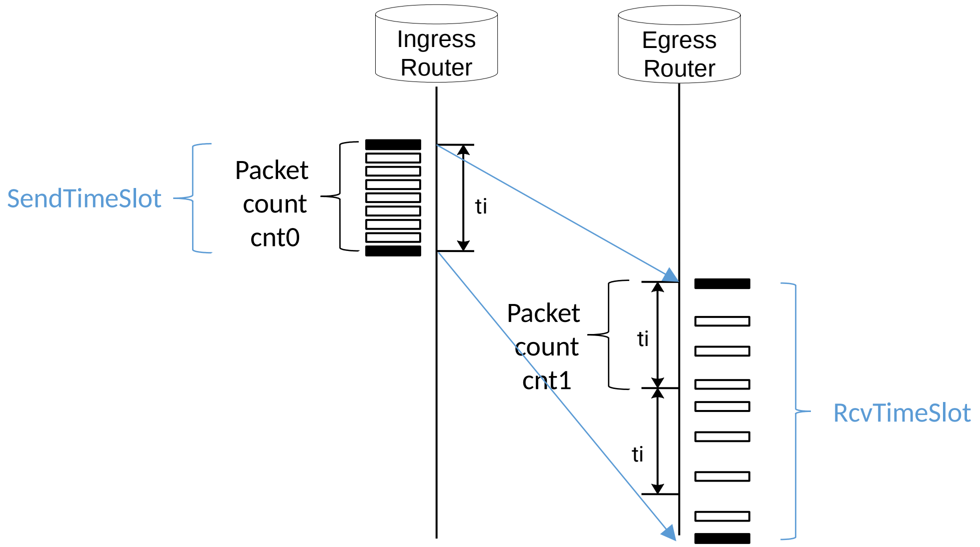
The Egress router read the cnt1 every ti interval, and send the result to the ingress; the Ingress gathers the results, and do calculation in everty ti*N interval. (e.g., ti=3.3ms, N=3)

- **TxRate** = (cnt0+cnt0'+cnt0''...) / ti*N
- **RxRate** = (cnt1+cnt1'+cnt1''...) / ti*N

PathCongestion = **RxRate** / **TxRate**

- The smallest one is the "worst" path; while the biggest one is the "best" path.
- If cnt<cnt0, it means there is packet loss happening, then the PathCongestion needs to be adjusted.

Key Technologies: Flow path switching - *Basic method*



$$\text{AftrSwitch\_PathCon} = \frac{(\text{cnt1}+\text{cnt1}'+\text{cnt1}''\ldots + \text{dRx} + \text{dTx}) \,/\, (\text{ti}*N + \text{CurPathJitter})}{(\text{cnt0}+\text{cnt0}'+\text{cnt0}''\ldots + \text{dRx} + \text{dTx}) \,/\, (\text{ti}*N)}$$

Other parameters:

- **CurPathJitter** = RcvTimeSlot-SendTimeSlot
- **dRx**: the count of flow(s) which is(are) planned to be switched into the current path
- **dTx**: the count of flow(s) which is(are) planned to be switched out of the current path

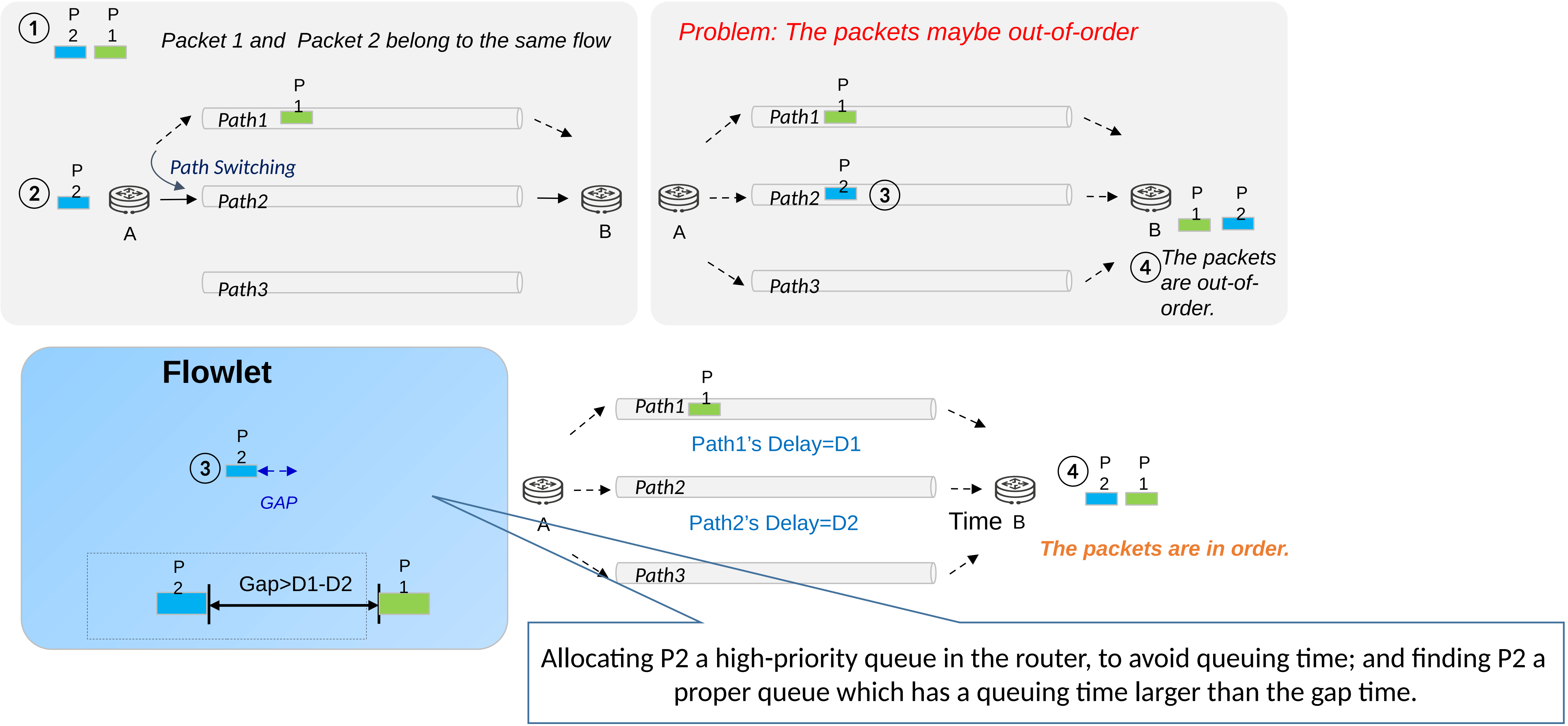
Basic Rules:

- Choose a flow in the “worst path”, and intend to switch it to the “best path”.
- Estimates the path congestion of each path, after the switching, according to the formula above. If the path congestion is more averaged than before, then the flow is considered a valid choice.
- Do the real path switch.
- Iterate above steps.

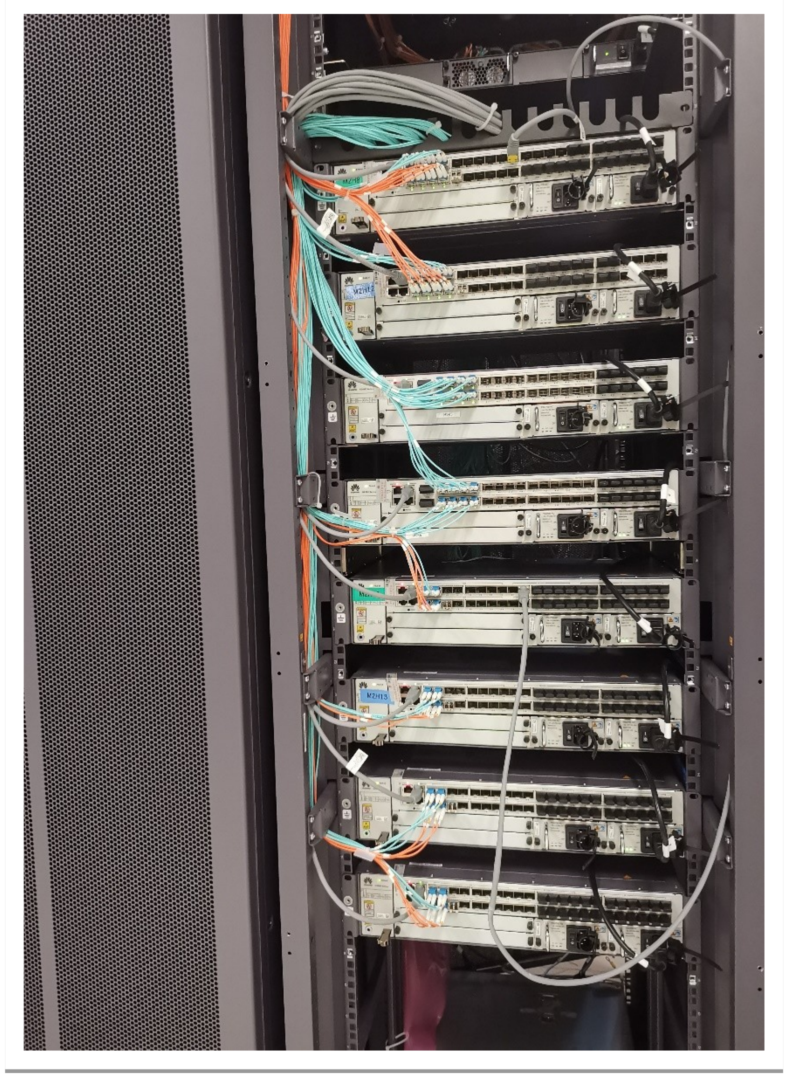
To avoid the flow switch oscillation, the flow that be switched would not be allowed to be switched again within a certain time slot (e.g. 5min).

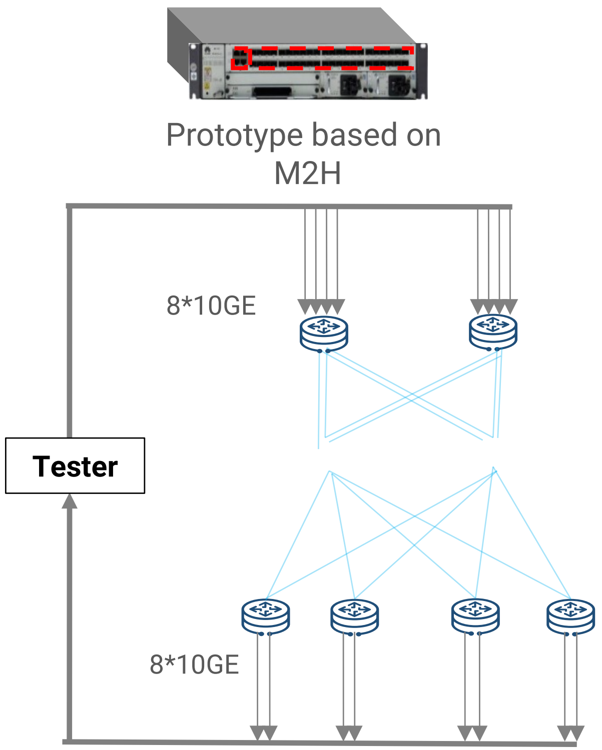
Key Technologies: Flow path switching – *Packet order assurance*

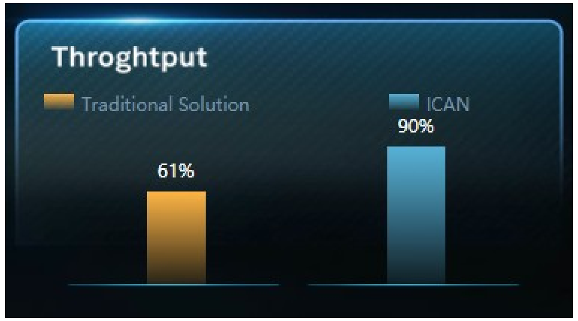
Flowlet-based Scheduling ensure no packet ordering/loss issue during path switching



Test Result







Premise:

- Network should support 80Gbps
- But at 48-50Gbps already started to drop packets
- Using iCAN, we improved this by around 30%
- Meaning network fills up more before signs of congestion

Discussion: Related standardization work

- iCAN architecture

  a) Traffic-engineering architectures for generic applicability across packet and non-packet networks. This includes, for example, net works that perform centralized computation and control, distributed computation and control, or even a hybrid approach. (from TE AS charter)

  - We believe that iCAN is a kind of “Traffic-engineering architectures ”
    - But not to replace current TE architecture, rather, to augment it
    - A hybrid approach, containing traditional centralized path planning/reserve and real-time path switching running in data plane
  - Aiming to define the architecture and technical requirements

- Path congestion metric

  b) Definition of protocol-independent metrics and parameters (measurement and/or service attributes) for describing links and tun nels/paths required for traffic engineering (and related routing, signaling and path computation). (from TEAS charter)

  - In conjunction with IPPM
    - https://tools.ietf.org/html/draft-dang-ippm-congestion-02
    - https://tools.ietf.org/html/draft-dang-ippm-multiple-path-measurement-02

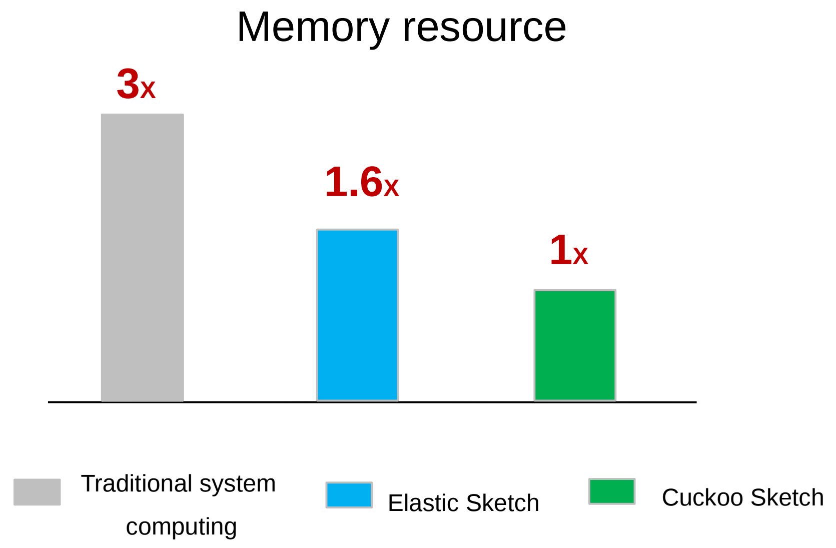
- Path quality assessment protocol
  - GRASP Extension (Anima WG)?
  - IPPM New Work (IPPM WG)?
  - Others?

Comments are appreciated very much!

Thank you!

IETF106, Singapore

Flow statistics within router



The CAIDA Anonymized Internet Traces (177K streams, 2M packets, maximum stream 16K packets)		
Algorithm	**Accuracy**	**Memory resource**
Traditional system computing	100%	~1MB
Elastic Sketch (SIGCOMM 2018)	≥99%	600KB
Cuckoo Sketch	≥99%	385KB

Enhanced Cuckoo Sketch Algorithm