Design for Large-Scale Collection System Using Flow Mediators

Atsushi Kobayashi, Tsuyoshi Kondoh, and
Keisuke Ishibashi
NTT Information Sharing Laboratories

Outline

- Introduction
  - Why do we need a large-scale collection system?
  - What is Flow Mediator?
- Requirements
  - I tried to explore the possibility of a large-scale collection system for large networks.
- Heuristic method of designing traffic collection system
  - Estimate number of flow records after aggregation or sampling
  - Adjust several parameters based on this result
- Summary

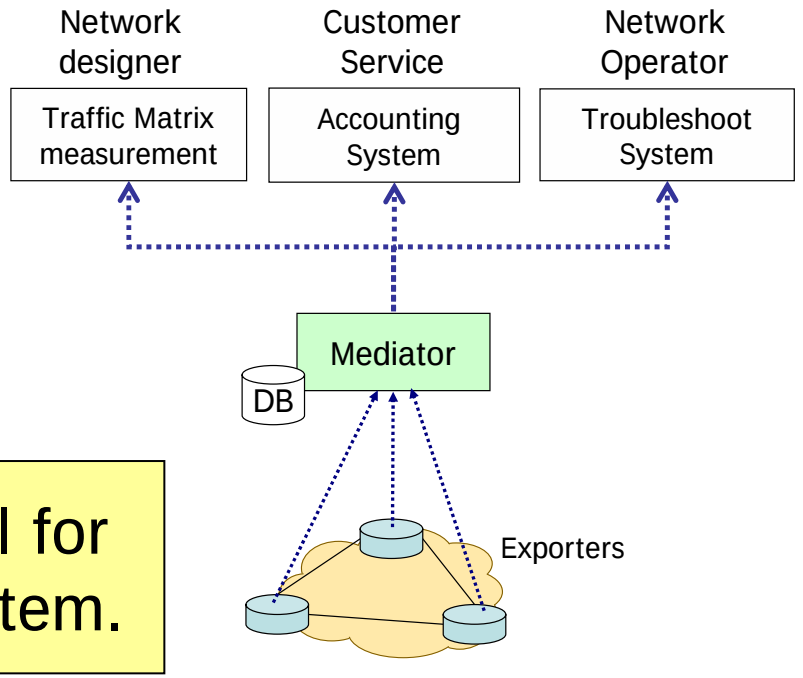
Introduction

- Traffic volumes in ISP networks are becoming huge in the last few years.
  - The number of exported flow records is becoming so huge that a single collector cannot handle them.
- A smaller sampling rate makes small flows invisible.
  - Even if traffic grows, network operators would like to maintain the same sampling rate as much as possible.
- Aggregated flow records from router make port number or IP address invisible.
  - Exporting 5-tuple flow records from router is better.

The demand for a large-scale traffic-collection system is growing.

What is Flow Mediator?

- Flow Mediator† is a device that “mediates” flow records and has the following functions:
  - collects Flow Records from various exporters
  - stores original flow records
  - aggregates flow records flexibly
  - distributes appropriate flow records for collectors/analyzers

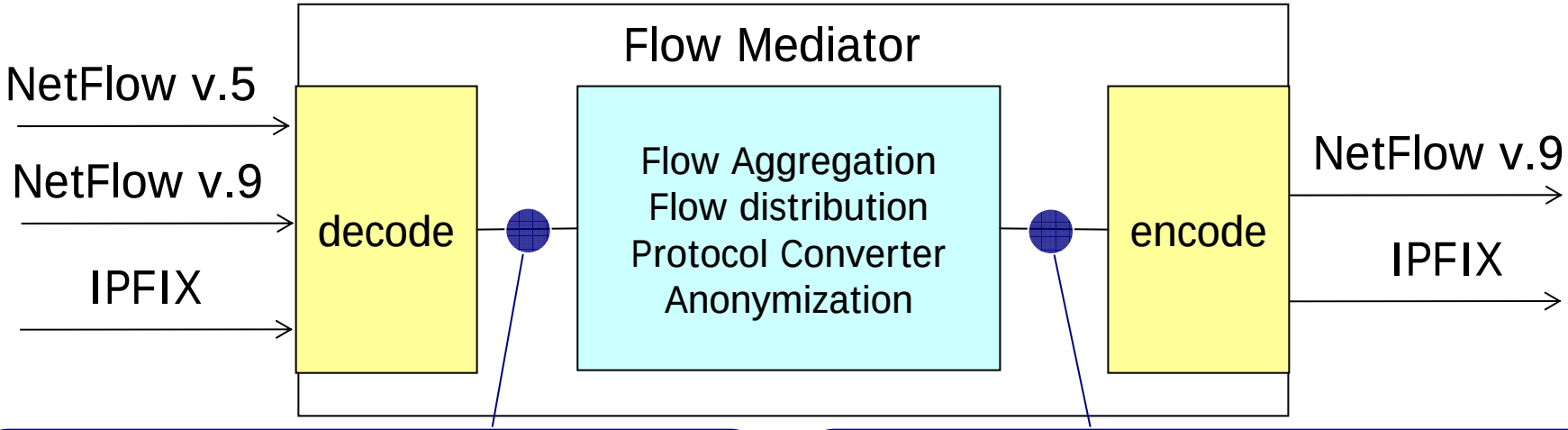


Flow mediator ought to be useful for making large-scale collection system.

† draft-kobayashi-ipfix-mediator-model-01.txt

You can easily make Flow Mediation code

- Net::Flow perl module is available on CPAN.
  - http://search.cpan.org/~akoba/Net-Flow-0.02/
  - The module can encode and decode NetFlow/IPFIX packets.
  - The encoding and decoding functions have a similar IF.




```
my ( $HeaderHashRef,
     $TemplateArrayRef,
   $FlowArrayRef,
   $ErrorsArrayRef) =
   Net::Flow::decode(
                  \$packet,
                  $TemplateArrayRef  ) ;
```

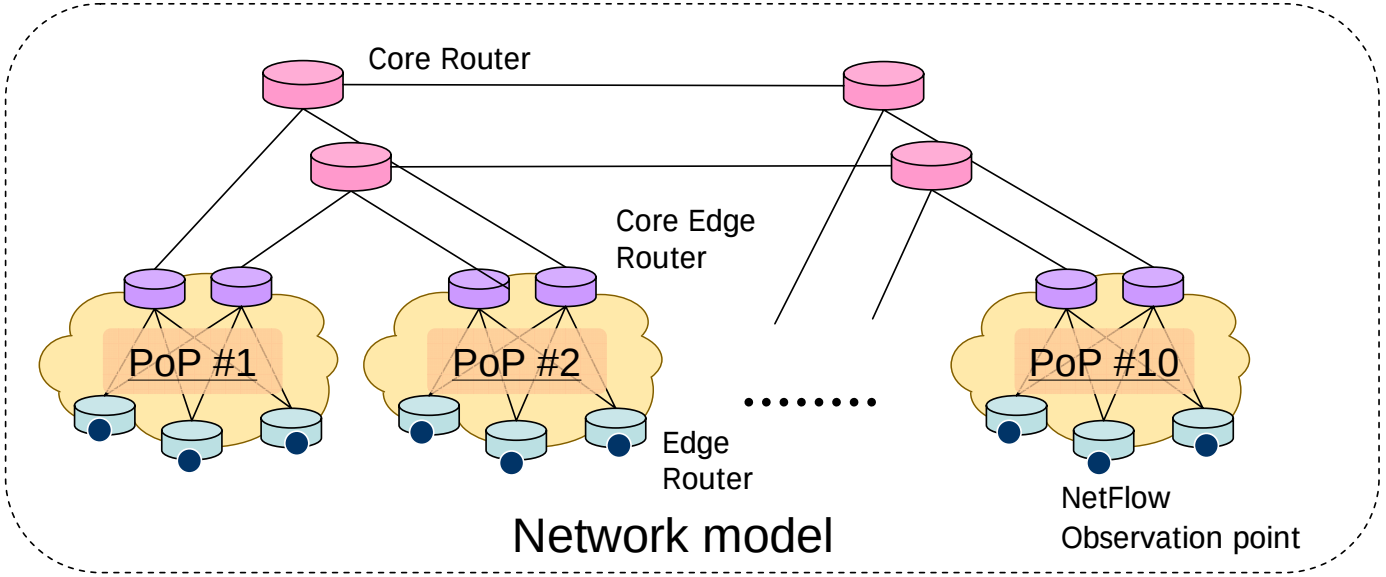
```
my ( $EncodeHeaderHashRef,
     $PktsArrayRef,
     $ErrorsArrayRef) =
     Net::Flow::encode(
                  $EncodeHeaderHashRef,
                  \@MyTemplates,
                  $FlowArrayRef,
                  1400 ) ;
```

Requirements

- Make traffic-collection system to meet following requirements
  - Requirement 1: measure traffic flow of entire networks
    - measure traffic matrices PoP by PoP and router by router
  - Requirement 2: store received 5-tuple flow records from router
    - When traffic incident happens, allow inspection of traffic.
  - Requirement 3: design scalable architecture to accommodate large ISP traffic volume

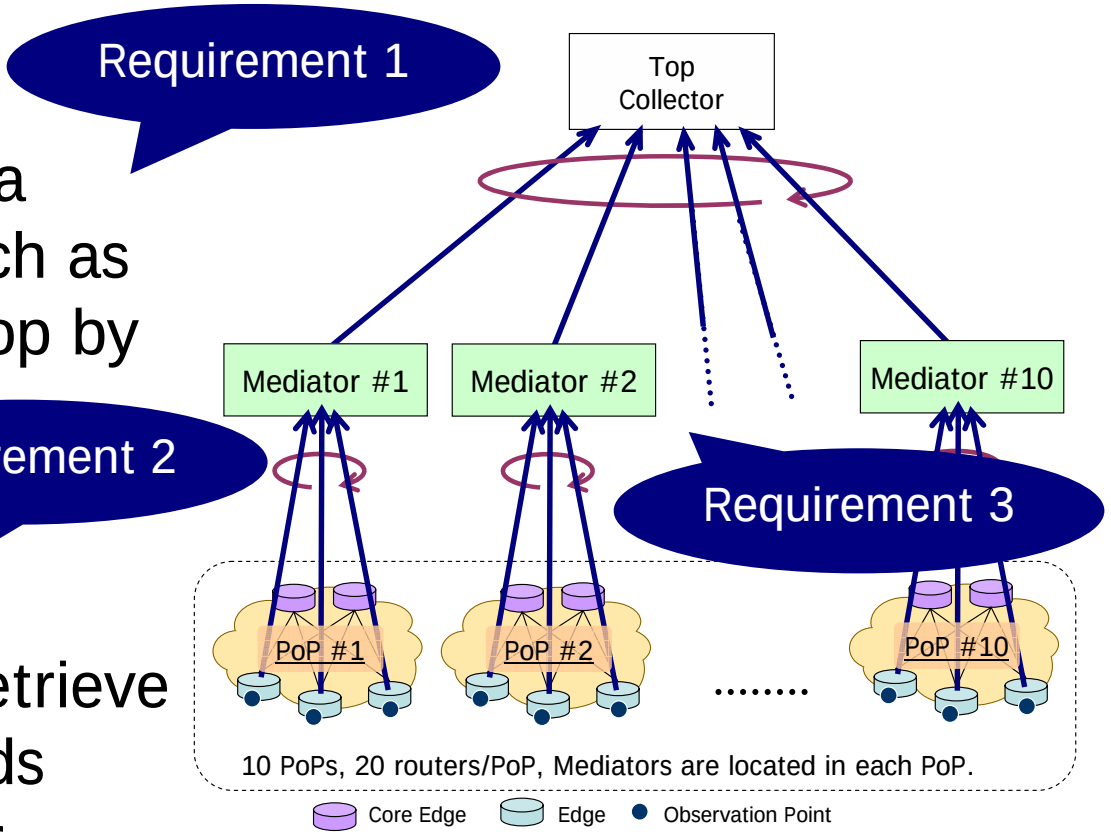
Goal

- Explore heuristic method of designing collection system for introduction into actual network.
- Proposed collection system needs to accommodate following network model.
  - Total traffic volume 500 Gb/s, 100 Mp/s
    - Edge Router 20/PoP×10 PoP = 200
    - NetFlow is enabled on IngressIF of Edge router.



Hierarchical Collection System

- Mediators are allocated in each PoP.
  - They store all flow records, aggregate them, and export them to next collector.
- Top Collector
  - measures wide-area traffic matrices, such as router by router, pop by pop.
- Inspection
  - If traffic incident happens, we can retrieve detailed flow records from Flow Mediator.

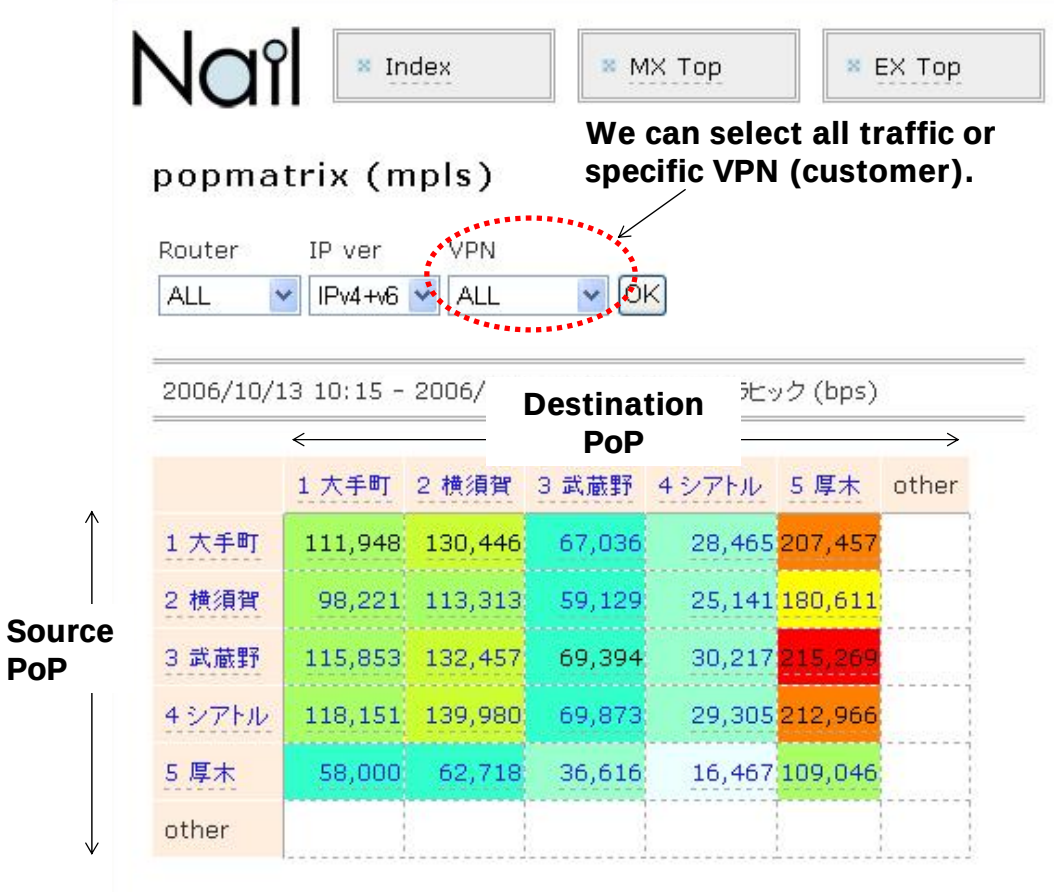


Visualize Traffic Matrices

- Top collector can visualize Router/PoP/AS Traffic Matrixes.

Nail is the name of our traffic matrix visualizer.

Color indicates traffic volume of Source/ Destination pair.



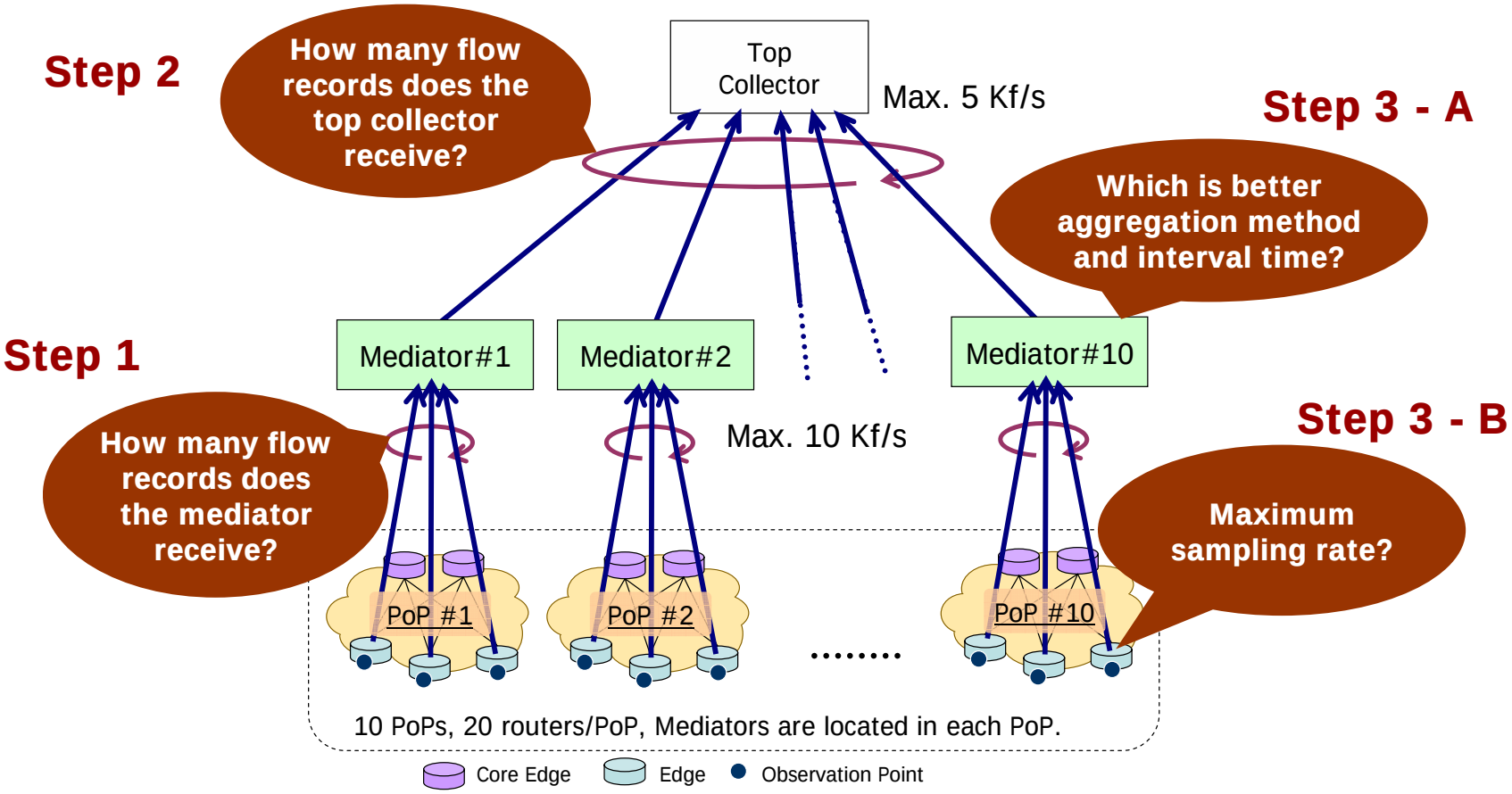
	1 大手町	2 横須賀	3 武蔵野	4 シアトル	5 厚木	other
1 大手町	111,948	130,446	67,036	28,465	207,457	
2 横須賀	98,221	113,313	59,129	25,141	180,611	
3 武蔵野	115,853	132,457	69,394	30,217	215,269	
4 シアトル	118,151	139,980	69,873	29,305	212,966	
5 厚木	58,000	62,718	36,616	16,467	109,046	
other						

Heuristic Design Method

- Suitable values of several parameters are decided by the following steps.
  - Step 0: measure performance limit of flow mediator and top collector.
  - Step 1: reveal relation between number of flow records and packet sampling
  - Step 2: reveal relation between number of flow records and aggregation that depends on several factors.
    - Aggregation methods (BGP Next-Hop, Prefix, host)
    - Aggregation interval time (20 s, 60 s, 90 s...)
  - Step 3: select suitable value within performance limit.
    - Large sampling rate is preferable.
    - Small granularity of aggregation is preferable.

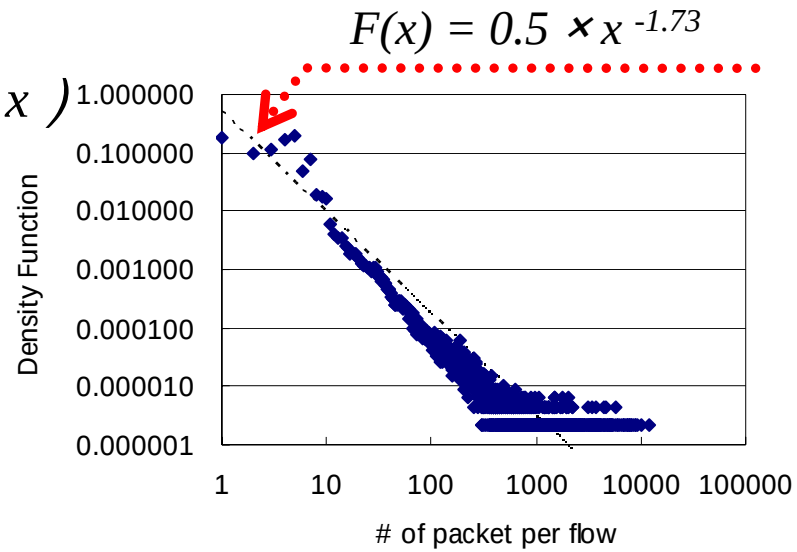
Consideration Points

- List several considerations, as follows.
  - Maximum performances of the top collector and mediators are 5 Kf/s and 10 Kf/s.



Step 1: estimate flow records after sampling

- Estimate number of flow records based on density function of packets per flow .
  - # of packets per flow: x
  - Packets per flow density function: $F(x)$
  - Sampling rate: $1/r$
  - Total number of unsampled flow: f_{all}



$$f_{sampled} = \sum_{x=1}^{\infty} \left(1-(1-1/r)^x\right) \times F(x) \times f_{all}$$

Extraction probability

$0.5x^{-1.73}$

Roughly estimate as follows.
100 Mpps ÷ 20 packets = 5 Mf/s

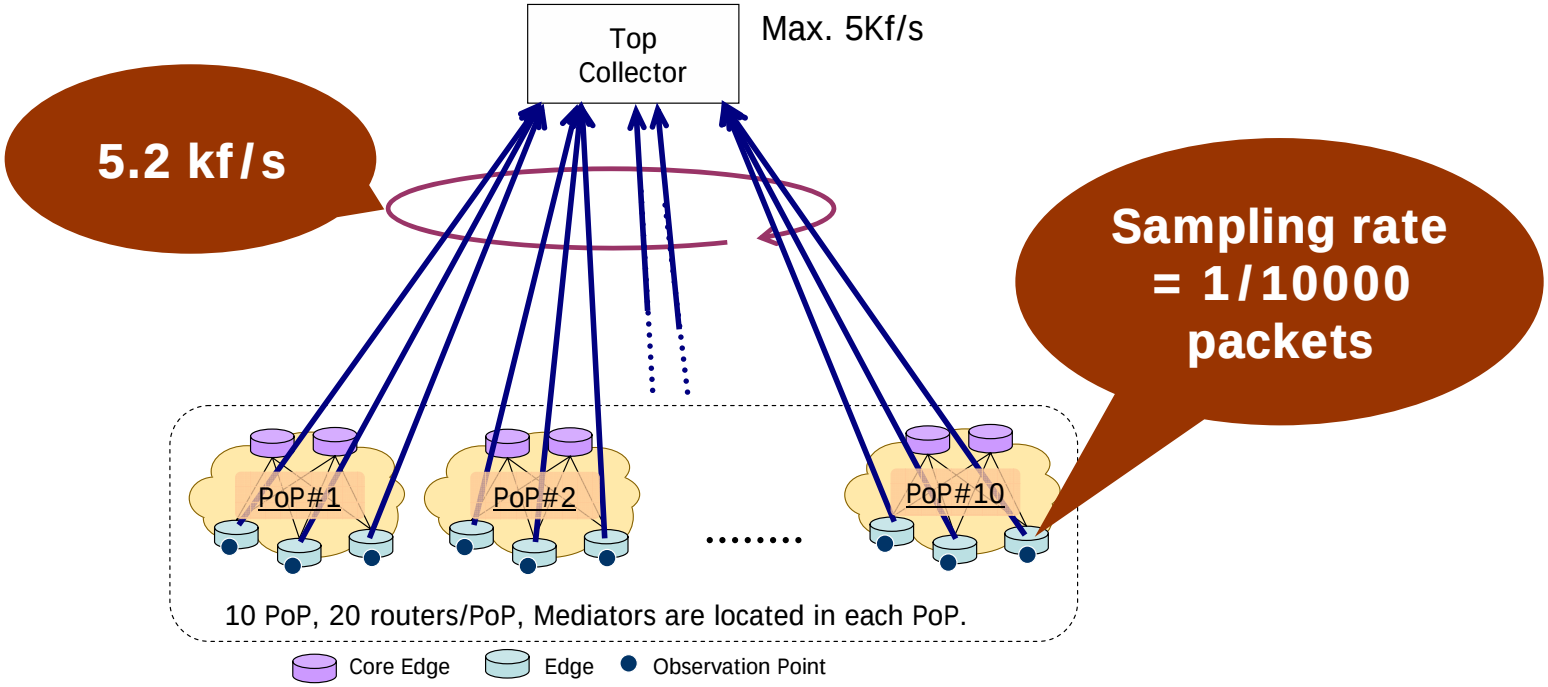
Approximate # of flows when total traffic volume is 500 Gb/s.

Sampling rate	1/100	1/1000	1/10000
$f_{sampled}$	*305 kf/s*	*43 kf/s*	*5.2 kf/s*

Too many flow records without mediator

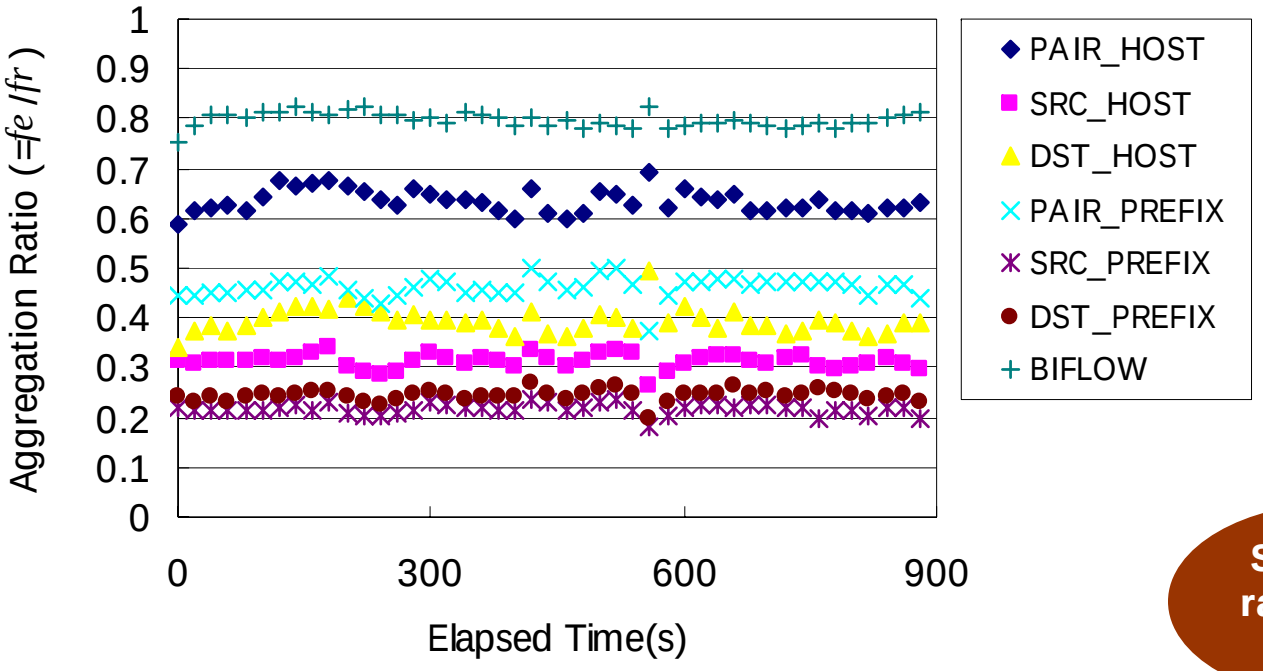
- Even if sampling rate is 1/10,000 packets, the number of flow records exceeds performance limit.

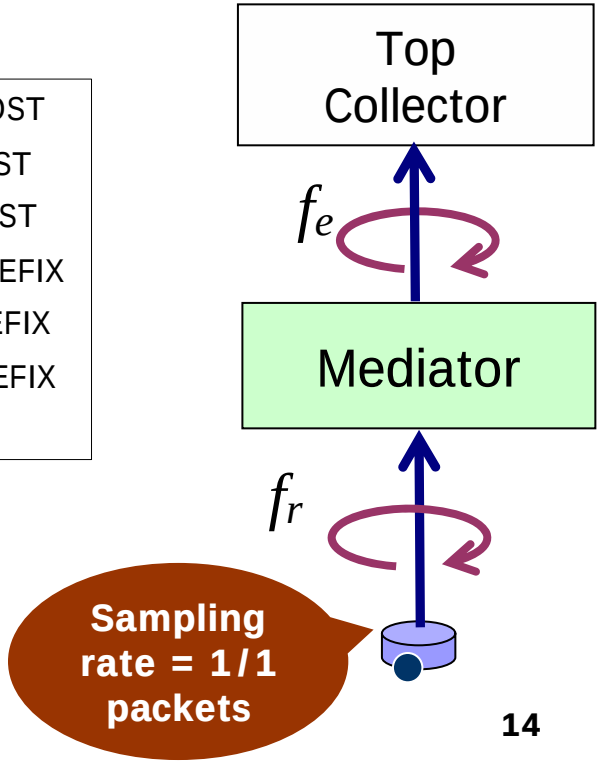
Sampling rate	1/100	1/1000	1/10000
$f_{sampled}$	*305 kf/s*	*43 kf/s*	*5.2 kf/s*



Step 2: flow records after aggregation

- What is the # of flow records after aggregation?
- Mediator aggregates unsampled flow records at 20-second interval.
  - Aggregation efficiency: Prefix > HOST > Pair Prefix > Pair HOST > Bi-Flow
    - The prefix length "/24" is uniformly applied to Prefix Aggregation.
    - Bi-flow is aggregated from two flow directions.

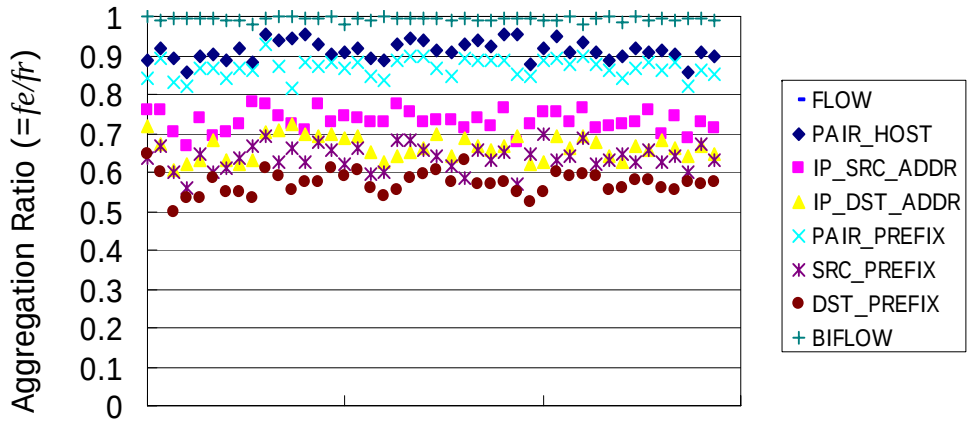




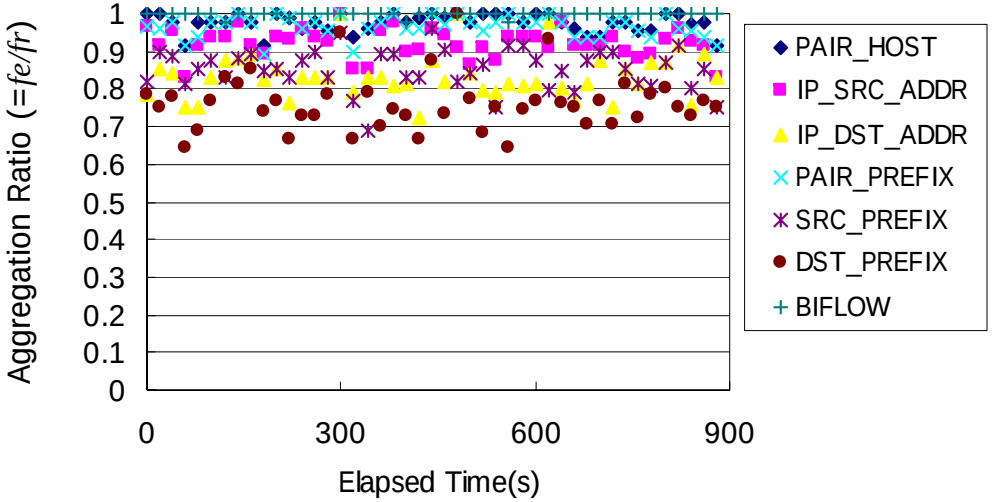
Step 2: Flow records after aggregation, sampling

- Each aggregation method becomes ineffective gradually.
- Bi-flow becomes ineffective immediately.
  - sensitive to sampling rate.

Sampling rate 1/128



Sampling rate 1/1024



Step 2: Which factor influences aggregation?

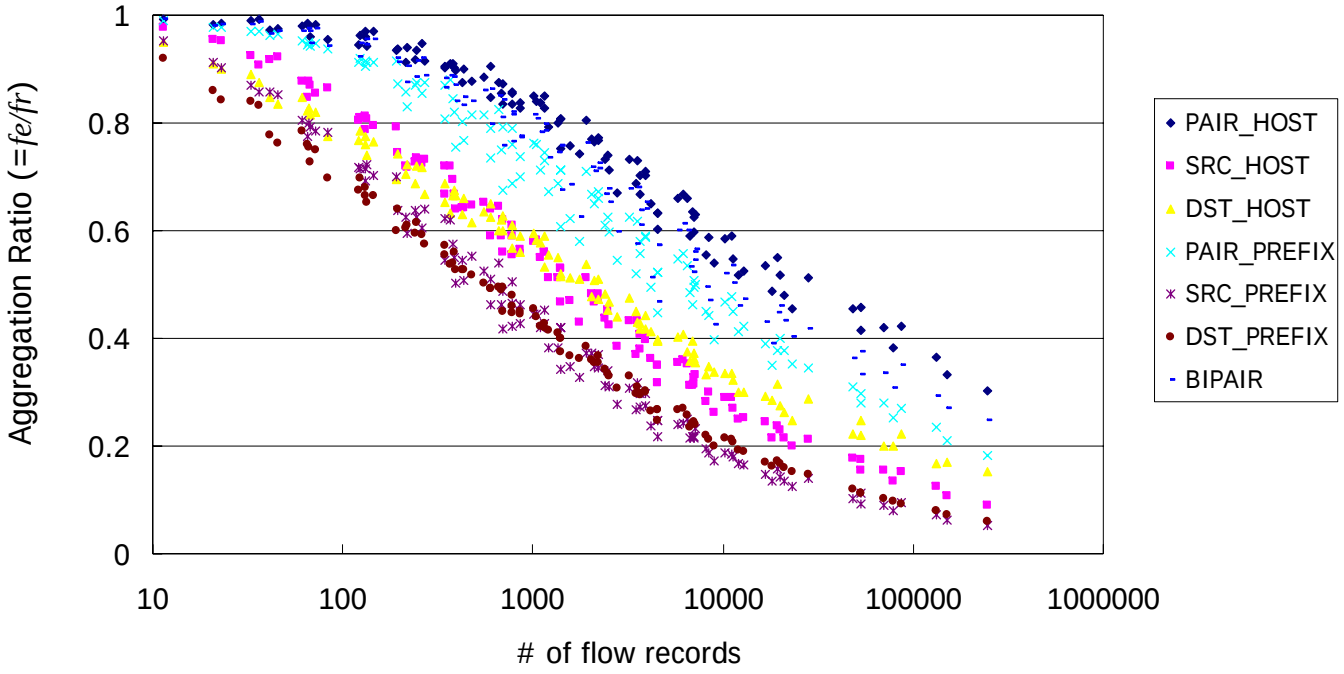
- Aggregation ratio depends on several factors.
  - Traffic Volume through observation point.
  - Sampling rate
  - Aggregation interval time

I guess that the aggregation ratio depends on the number of flow records received in interval time.

Received Flows	3450	3562
Aggregation Interval Time (s)	10	300
Sampling rate (1/r)	1	128
DST_HOST Aggregation ratio	45%	43%
DST_PREFIX Aggregation ratio	30%	32%

Step 2: Which factor influences aggregation?

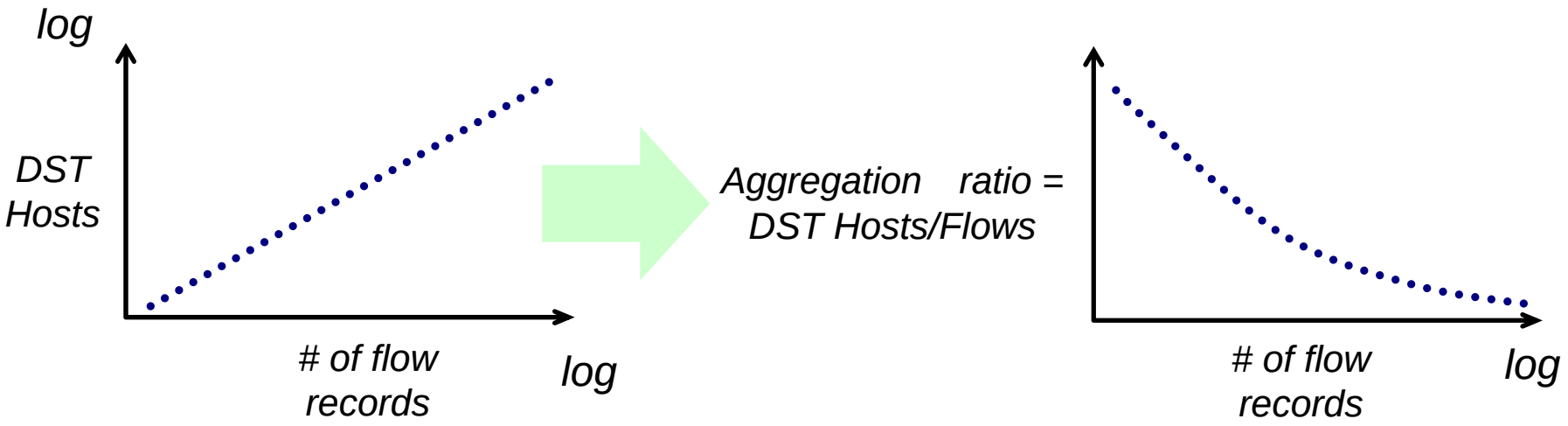
- I plotted all experimental data into one graph.
  - Three MAWI traffic data samples have different volumes.
  - Aggregation Interval time : 5 – 300s
  - Sampling rate : 1/1 – 1/1024



Aggregation ratio depends on number of received flow records.

Step 2: Formulation of Aggregation Ratio

- Aggregation ratio (R) can be estimated from number of flow records (f_r), as follows.
  - DST Host aggregation: $R_{dsthost} = 1.80 \times f_r^{-0.18}$
  - DST Prefix aggregation: $R_{dstprefix} = 2.34 \times f_r^{-0.26}$
- After all, the aggregation ratio depends on the # of unique hosts or prefixes versus # of flows.



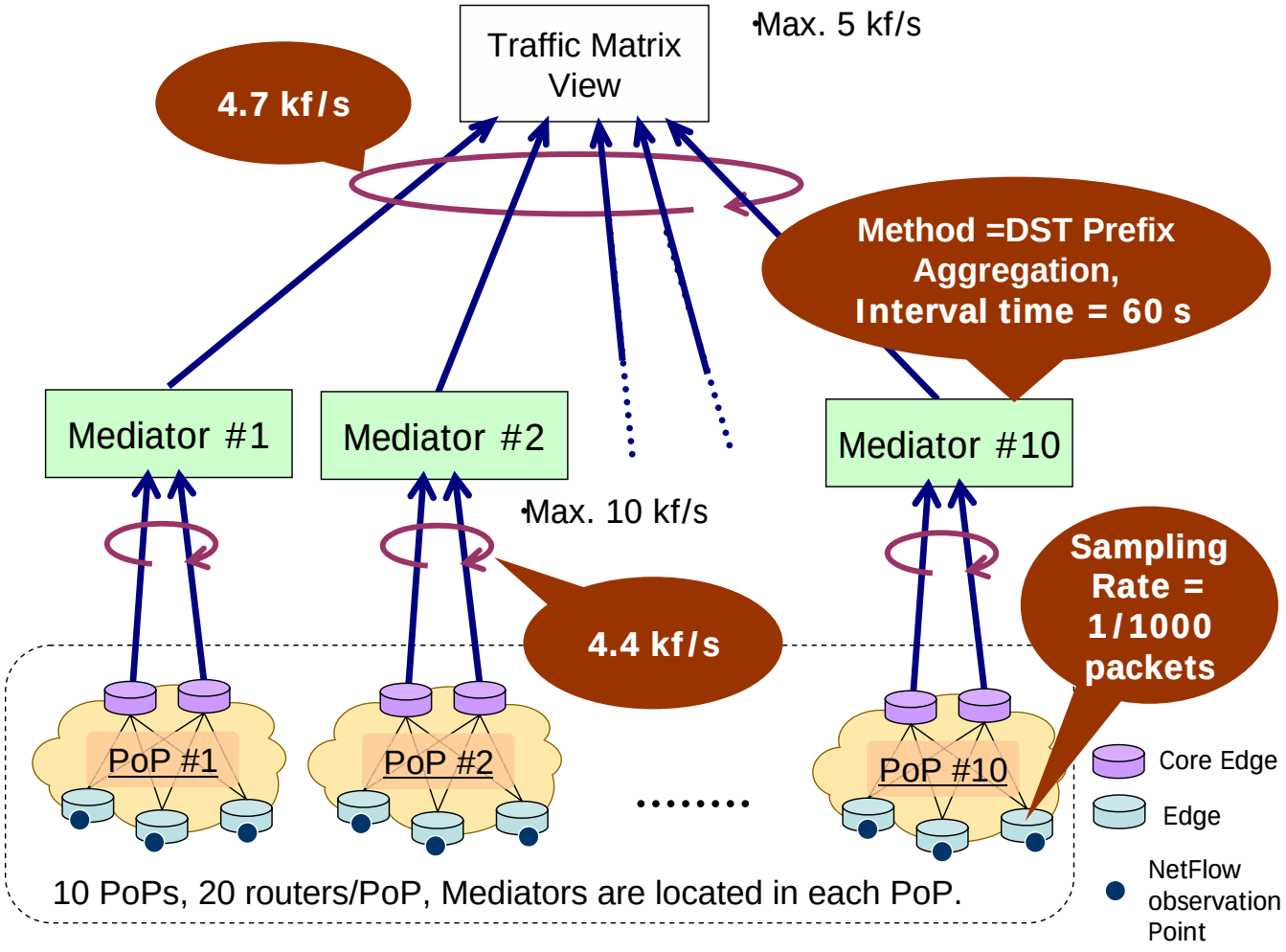
Step 3: Selection of Suitable Values

- I selected suitable value within performance limit.

		Sampling Rate	1/100	1/1000	1/10000
# of received flow records in top collector ($=\sum f_e$)	DST_HOST aggregation	Interval time = 60s	*45 kf/s*	9.0 kf/s	***1.6 kf/s***
	DST_Prefix aggregation	Interval time = 60s	*21 kf/s*	***4.7 kf/s***	***0.94 kf/s***
	DST_HOST aggregation	Interval time = 300s	*34 kf/s*	7.0 kf/s	***1.2 kf/s***
	DST_Prefix aggregation	Interval time = 300s	*12 kf/s*	***3.0 kf/s***	***0.62 kf/s***
# of received flow records in mediator (f_r)	—	—	*30 kf/s*	***4.4 kf/s***	***0.6 kf/s***

Example of collection system

- Sampling Rate: 1/1000
- Aggregation Interval time: 60 s



Conclusion

- To make large scale traffic collection system, flow mediator is efficient.
- Revealed relation between number of flow records and several factors:
  - Traffic volume
  - Sampling rate
  - Aggregation method
  - Aggregation interval time
- Demonstrated that traffic collection system using mediator can be introduced into actual large-scale networks.

Thank you for your attention.

This study was supported by the Ministry of Internal Affairs and Communications of Japan.