ICNP 2021

Diagnosis-driven SLO Violation DEtection

Yiran Lei, Yu Zhou, Yunsenxiao Lin, Mingwei Xu, Yangyang Wang

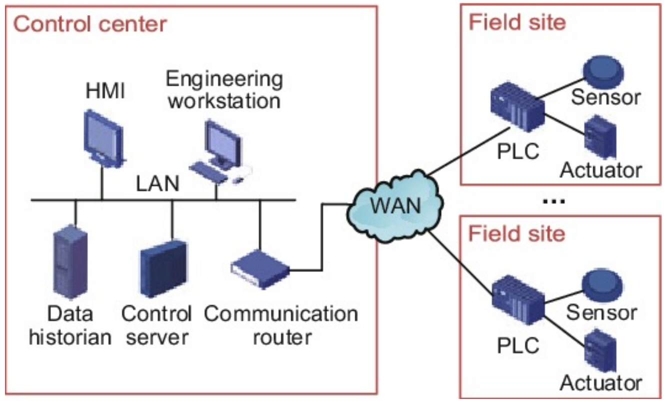
Tsinghua University

SLO Maintenance is Important

Algorithmic Stock Trading

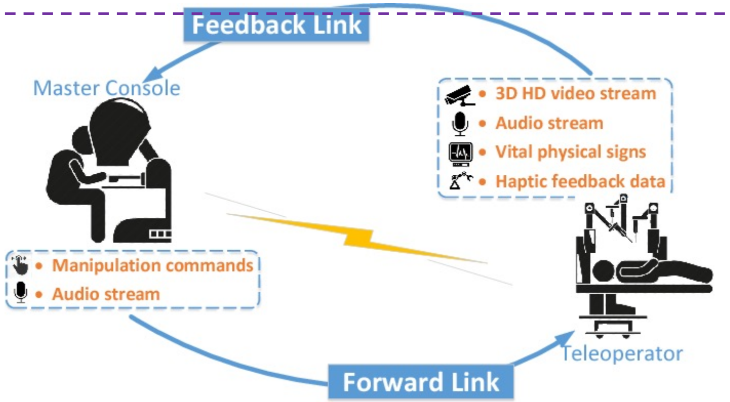
Cloud Gaming

✓ constantly low latency



Industrial Control System

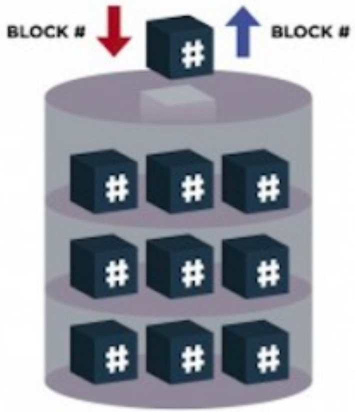
✓ constantly low latency and packet loss



Telesurgery

Video Conference

✓ constantly low latency and packet loss, and high throughput



Block Storage Service

VR Video Streaming

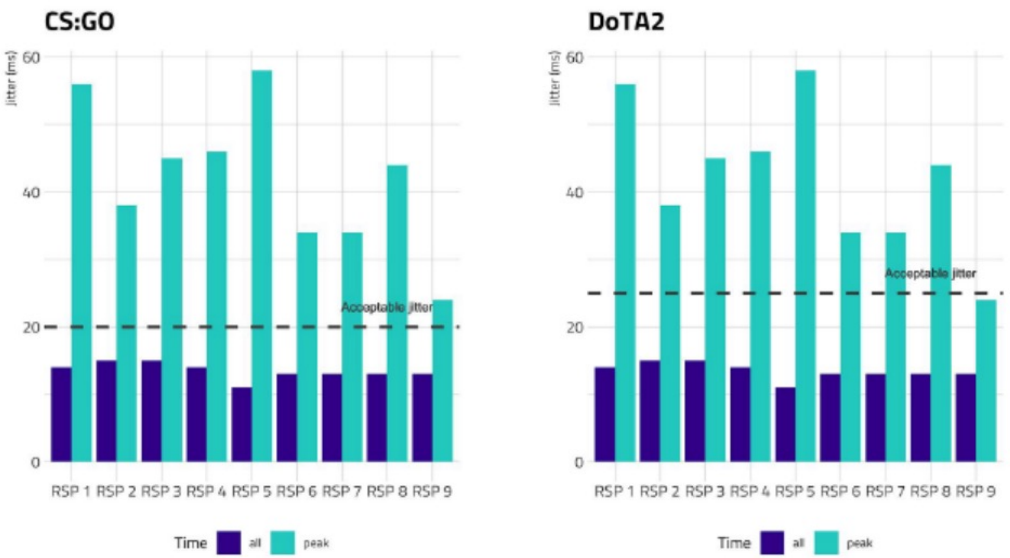
✓ high throughput and low packet loss

SLO Violation is Common and Destructive

	Distinct Queries/User	Query Refinement	Revenue/User	Any Clicks	Satisfaction	Time to Click (increase in ms)
50ms	-	-	-	-	-	-
200ms	-	-	-	-0.3%	-0.4%	500
500ms	-	-0.6%	-1.2%	-1.0%	-0.9%	1200
1000ms	-0.7%	-0.9%	-2.8%	-1.9%	-1.6%	1900
2000ms	-1.8%	-2.1%	-4.3%	-4.4%	-3.8%	3100

- Means no statistically significant change

Latency SLO violation causes monetary damages from Google and Bing

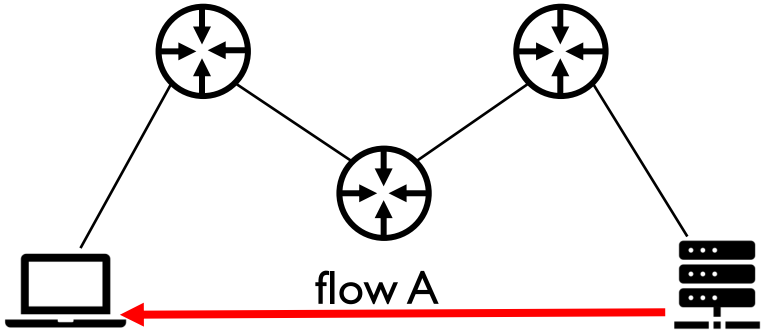


Latency SLO violation is common at busy time when playing games from Australian ISPs

- Amazon lost $66,240/ minute on 2013.8.19 due to a blackout
- 40-80 machines suffer from packet loss in DCN per year
- Katz-Bassett discovered reachability problems involving about 10,000 distinct prefixes during 3 weeks
- Tens of Internet outages from https://www.thousandeyes.com/outages/ in last 24 hours

Fast Mitigation upon SLO Violation

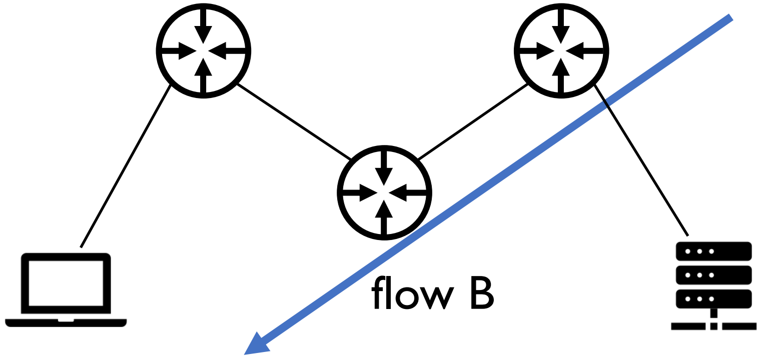
detection → diagnosis → troubleshooting



- delay of flow A exceeds 120ms
- delay SLO is violated

⇩

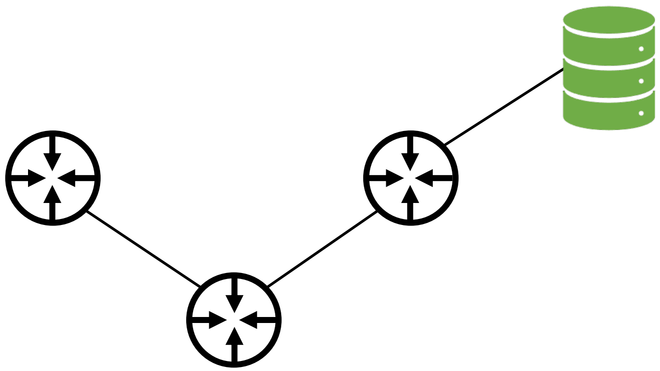
discover SLO violation:
✓ measure performance
✓ compare with objectives



- traffic of flow B bursts
- flow A's SLO violation is due to flow B's burst

⇩

analyze causality of SLO violation:
- find flow-level causes



- a server is sending large flow B
- the server is misconfigured
- fix and mitigate violation

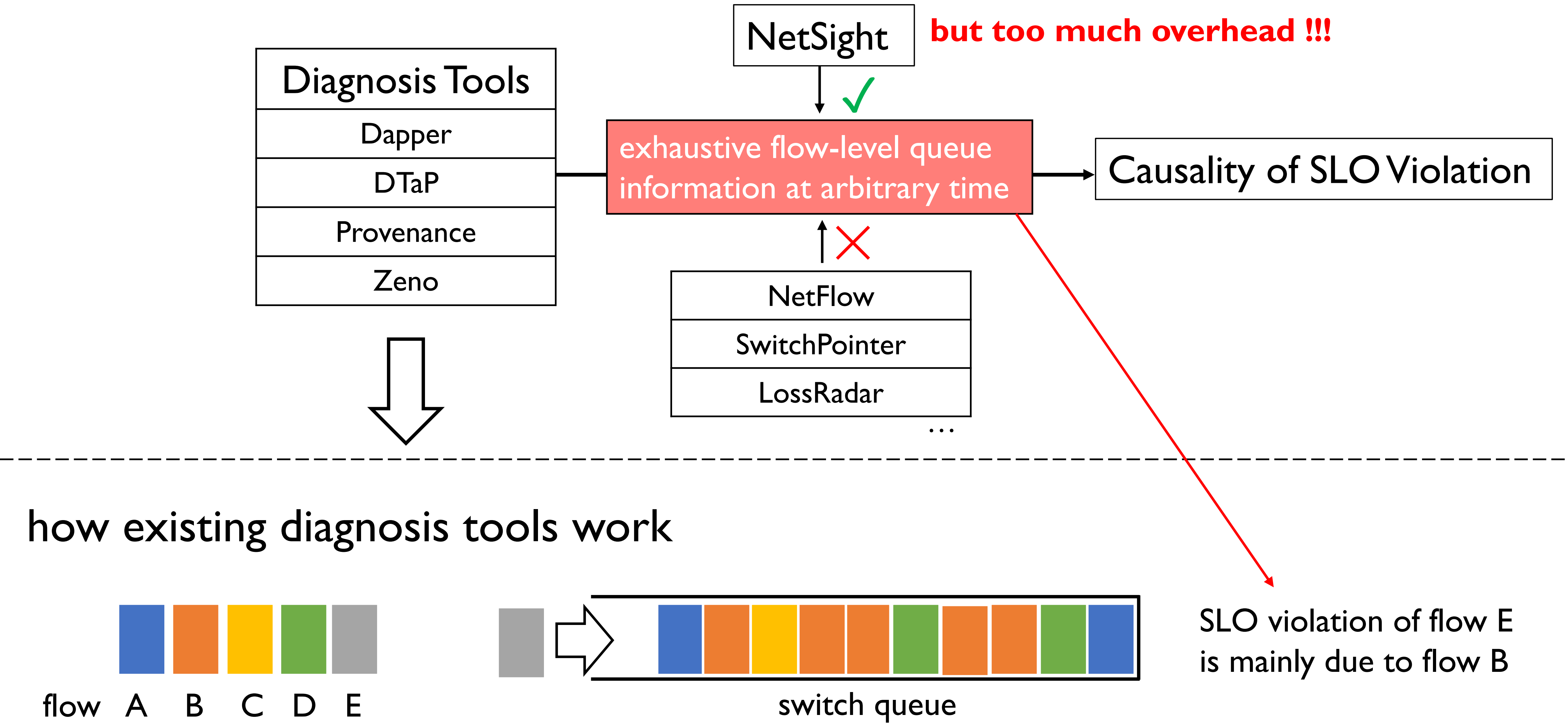
⇩

repair hardware and software
…

Problems of Existing Solutions

Detection Tools								
Existing Solutions	Property					SLO Type		
	granularity	lags	overhead	control plane involvement	end-host involvement	packet loss	percentile delay	max delay
ping	coarse	low	low	×	✓	✓	✓	✓
Netflow	coarse	high	low	×	×	×	×	×
SNMP	coarse	high	low	×	×	×	×	×
NetSight	fine	low	high	✓	×	✓	✓	✓
SwitchPointer	fine	low	low	×	✓	×	×	×
TPP	fine	low	high	×	✓	✓	✓	✓
LossRadar	fine	high	low	✓	×	✓	×	×
INTSight	fine	low	low	✓	×	×	×	✓
???	**fine**	**low**	**low**	×	×	✓	✓	✓

Problems of Existing Solutions



DOVE: Diagnosis-driven SLO Violation Detection

Detection								Diagnosis
Property					SLO Type			✓
granularity	lags	overhead	control plane involvement	end-host involvement	packet loss	percentile delay	max delay	
fine	**low**	**low**	×	×	✓	✓	✓	

Specific Flows

resource constraints of the data plane

detection

detection does not rely on the control plane

accurate SLO verification

high diagnosis accuracy: gather more causal information

diagnosis

selected flows

partial set of flows instead of the complete set of flows

empirical and set by network operators

watched flows

a wider set of flows, even the complete set of flows

ONLY data plane resources & logics

data plane logics & resources

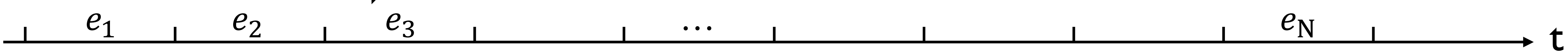
sketches

Sonata

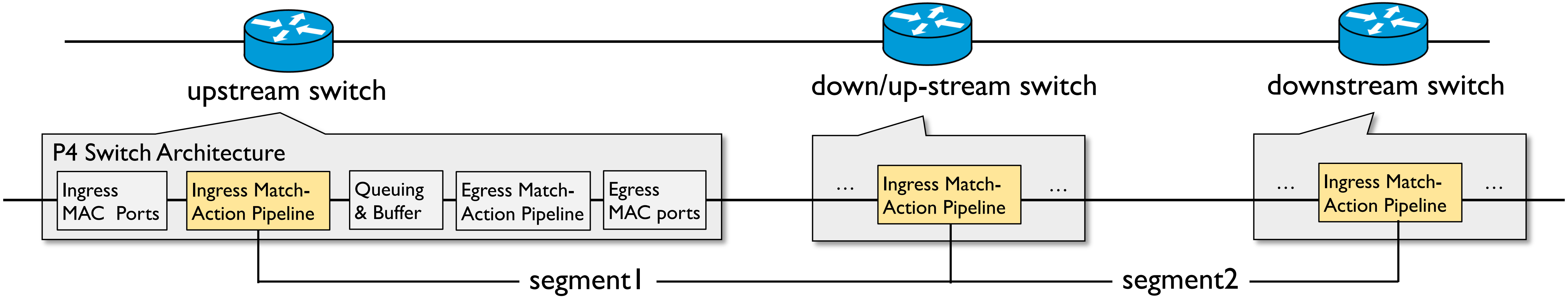
…

Epoch and Segmentation

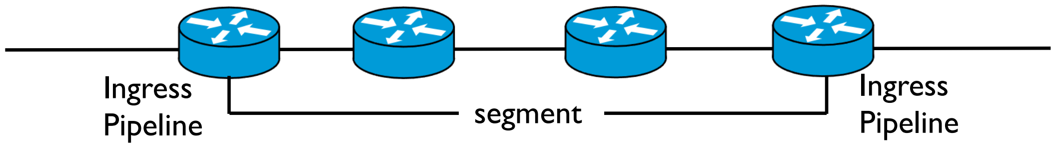
continuous time ⇨ adjacent epochs (hundreds of microseconds)



flow path ⇨ segments



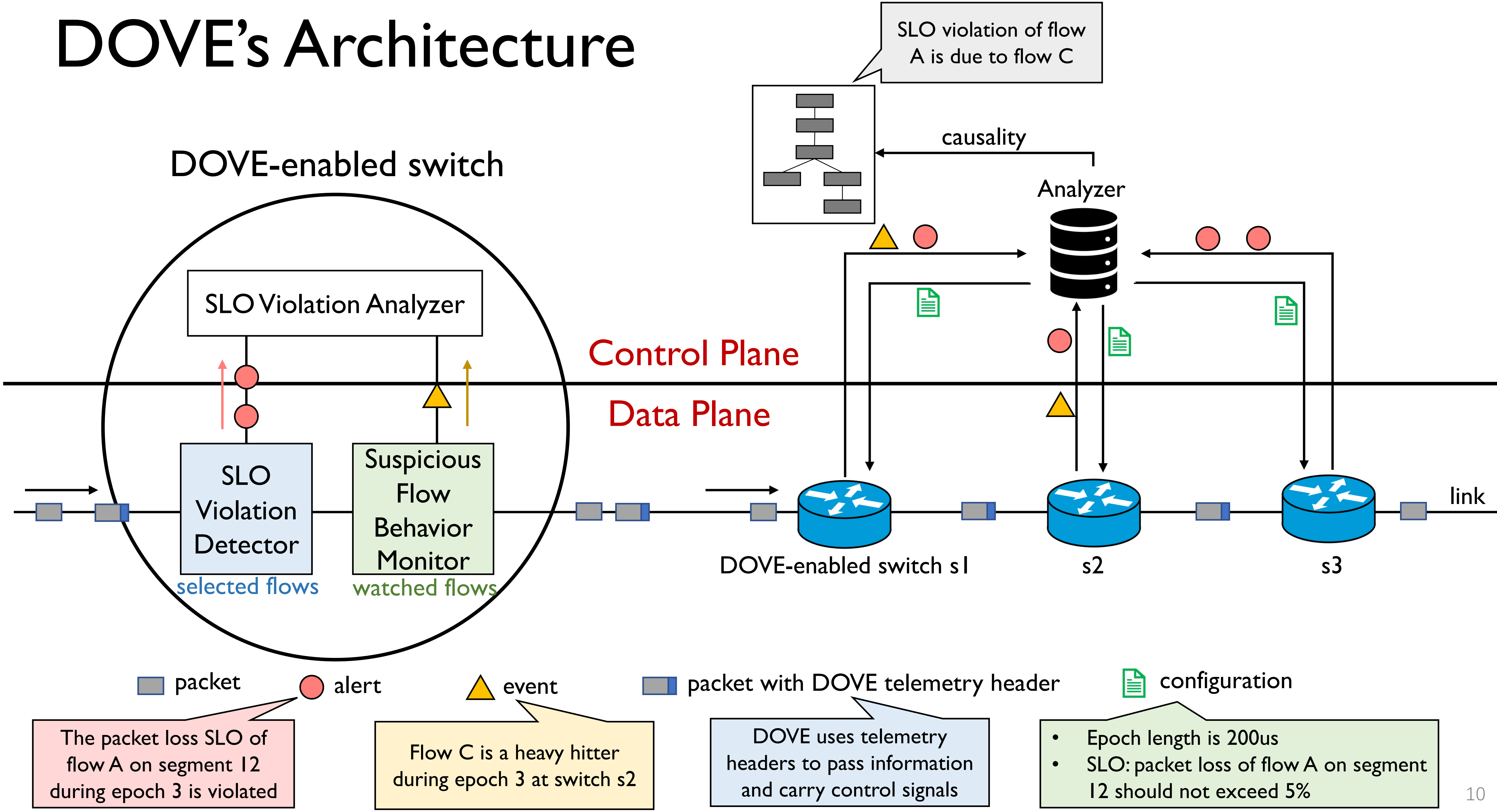
upstream and downstream switches ≠ neighbor switches



⇨ partial and incremental deployment

DOVE measures SLOs **on each segment during each epoch**

DOVE's Architecture

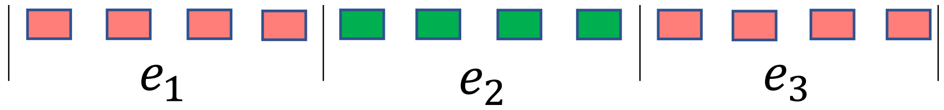


SLO Violation Detector

✓ Packet Loss

- Coloring Algorithm: the number of lost packet
- completely on the data plane

① upstream switch dyes packets red or green



② switch's red/green counter records

upstream ++ when switch sends

upstream ++ when switch sends

downstream ++ when switch receives

downstream ++ when switch receives

③ upstream switch copies the counter value of previous epoch into the DOVE telemetry header

carries the value of carries the value of

④ upstream switch clears control bit for the first half of epoch and sets the bit for the second

$t \in [0, e/2)$: carries control bit = 0

$t \in [e/2, e)$: carries control bit = 1

resets resets

⑤ downstream switch stores upstream counter value and calculates packet loss upon first set control bit

stores the value of stores the value of

control bit = 1: PL = - resets

PL = - resets

SLO Violation Detector

✓ Percentile Delay
- approximation algorithm
- on the data plane

difficult	measure flow-level percentile delay
	what is η-percentile delay of a series of measure delays?

verify SLO of percentile delay	feasible
does η-percentile delay exceeds threshold d?	

Given N values sorted in ascending order, the η-percentile value is:

I. $(1 + (N-1) \cdot \eta\%)$ -th sorted value, if $(N-1) \cdot \eta\%$ is an integer

II. some value between $1 + \lfloor (N-1) \cdot \eta\% \rfloor$ -th and $1 + \lceil (N-1) \cdot \eta\% \rceil$ -th value, if not

Statement1: η-percentile value $> d$

Statement2: let the number of value exceeding d be n, $n > N - \lfloor (N-1) \cdot \eta\% \rfloor - 1$

I. if $(N-1) \cdot \eta\%$ is an integer, Statement1 $\Leftrightarrow$ Statement2

proof: η-percentile value is $(1 + (N-1) \cdot \eta\%)$-th value

II. if not, Statement1 $\Leftarrow$ Statement2

proof: If $n = N - \lfloor (N-1) \cdot \eta\% \rfloor - 1$, d can be any value between $1 + \lfloor (N-1) \cdot \eta\% \rfloor$ -th and $1 + \lceil (N-1) \cdot \eta\% \rceil$ -th value. In this case, η-percentile value can have any size relations to d.

SLO Violation Detector

✓ Percentile Delay

Statement1: η-percentile value $> d$

Statement2: let the number of value exceeding d be n, $n > N - \lfloor (N-1) \cdot \eta\% \rfloor - 1$

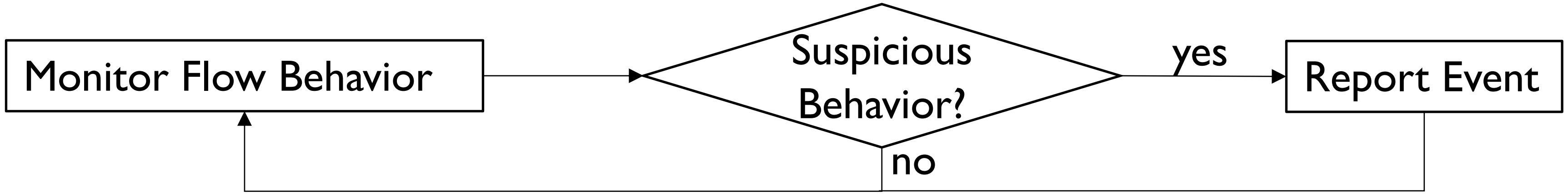
Statement1 $\xLeftrightarrow{approximation}$ Statement2:

calculate $N - \lfloor (N-1) \cdot \eta\% \rfloor$ from control plane and populate it as a threshold to the data plane

✓ Max Delay

compares the new measured delay with history delays and stores the bigger one

Suspicious Flow Behavior Monitor



Unlike SLO measuring, the monitor checks flow behaviors **on each DOVE switch during each epoch**

Suspicious Flow Behavior Monitor

what contributes to SLO violations ?

SLO Violation Causes

- high queue occupancy
- inter- and intra-switch loop
- …

high delay

- queue overflow
- link corruption and failure
- software bugs
- …

high packet loss

the most common one

queue-related cause

flows contributing much to queue occupancy

✓ Heavy Hitter:
monitor flows with large traffic

✓ Heavy Changer:
monitor flows whose traffic increases rapidly
monitor newly-established flows

SLO Violation Analyzer

principle:

1. location adjacency: flows sharing same queues
2. epoch adjacency: flows having close epochs

correlate the alert to:

1. high queue occupancy at upstream switch:
   - ☐ events from alert's upstream switch
   - ☐ events sharing same egress port with the alert
   - ☐ events happen just before alerts
   - ☐ any heavy hitters or heavy changers

2. high queue occupancy at the previous switch of the downstream switch:
   - ☐ events from alert's downstream switch
   - ☐ events sharing same ingress port with the alert
   - ☐ events happen just before alerts
   - ☐ any heavy hitters and heavy changers

alert
flow ID
upstream switch id
egress port
downstream switch id
ingress port
if violate max delay SLO
if violate percentile delay SLO
if violate packet loss SLO
epoch

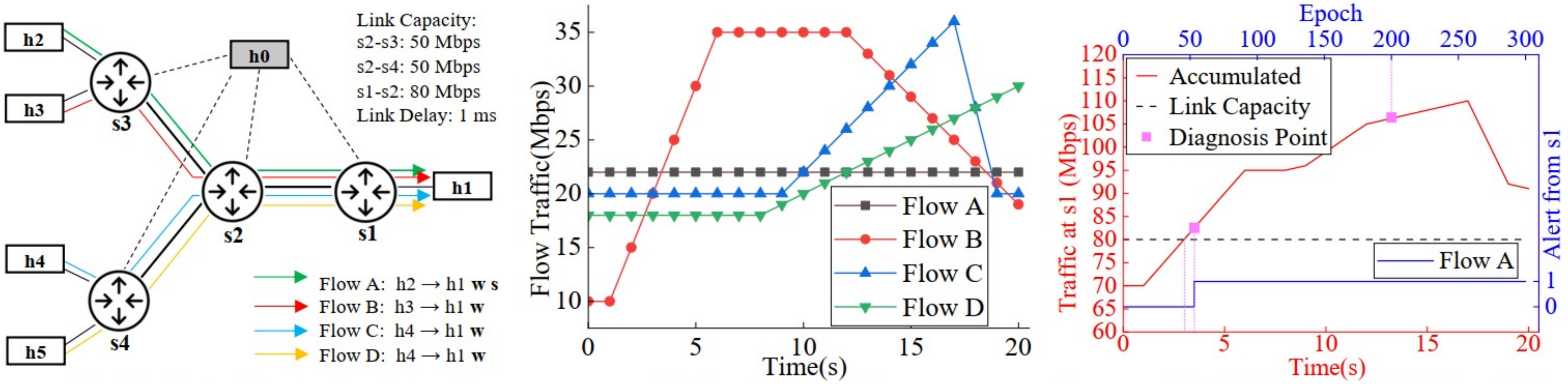
event
flow ID
switch id
ingress port
egress port
if heavy hitter
if heavy changer
epoch

Evaluation

Case Study - DOVE's effectiveness

settings and collected alerts:

diagnosis point 1 - 3.43s
diagnosis point 2 - 13.1s



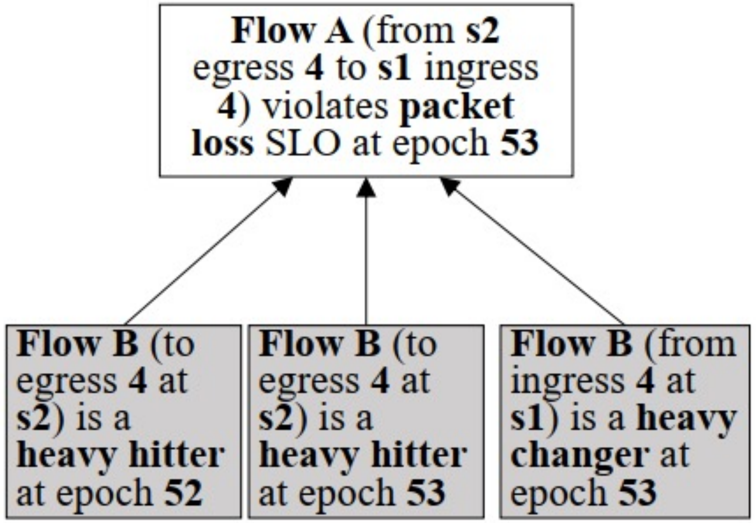
(a) network topology and routing (b) flow A,B,C,D rate (c) accumulated traffic on s2 → s1 link and flow A alerts received from s1

selected flow A suffers performance degradation from flow B,C,D competition on link s2-s1

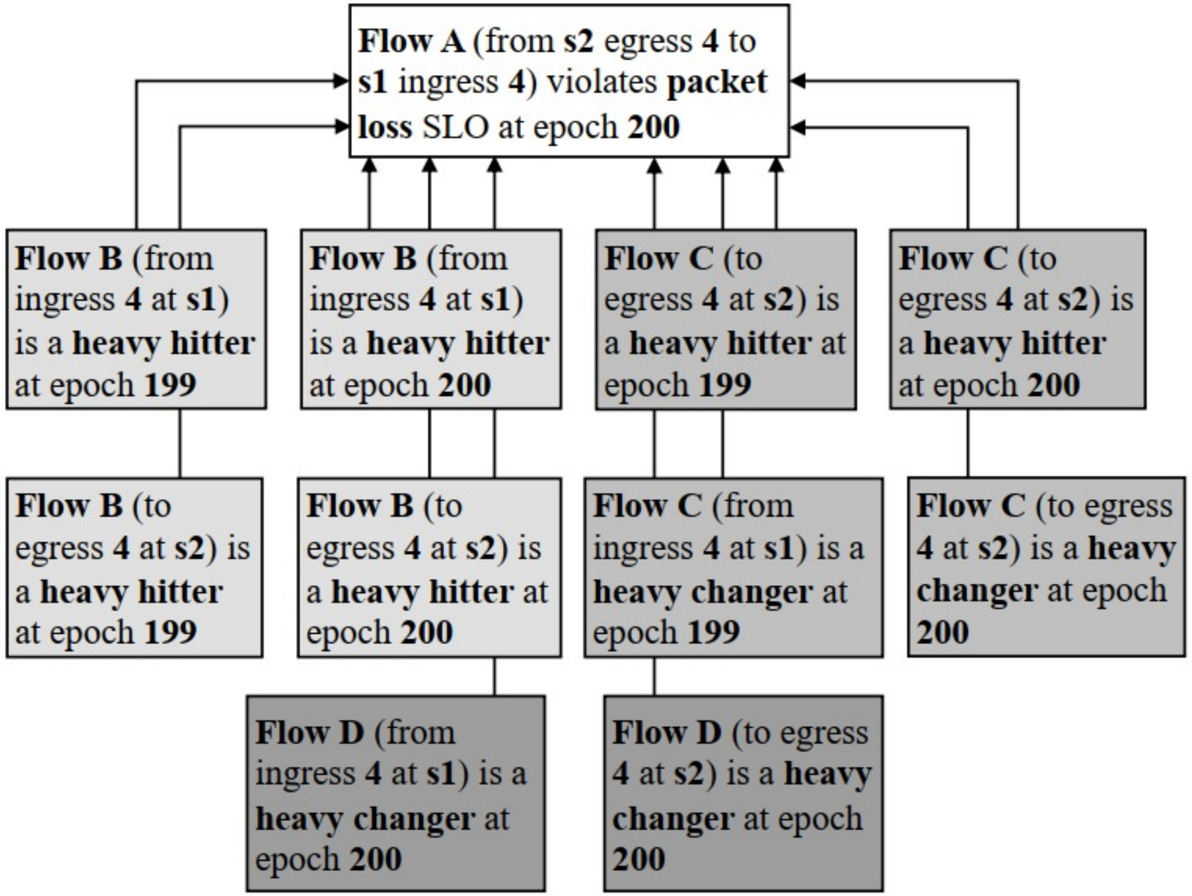
Evaluation

Case Study: DOVE's effectiveness

diagnosis results:



(a) provenance on diagnosis point 1

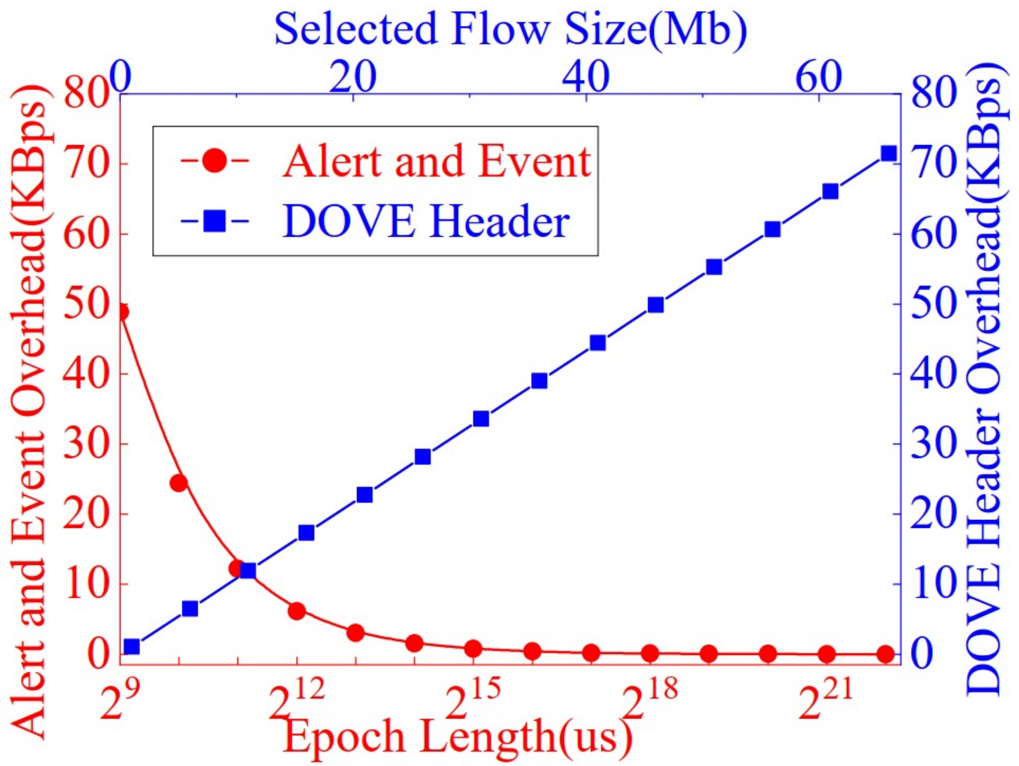


(b) provenance on diagnosis point 2

- diagnosis point 1:
  only flow B is the culprit flow
- diagnosis point 2:
  flow B, C, D are all the culprit flows

Evaluation

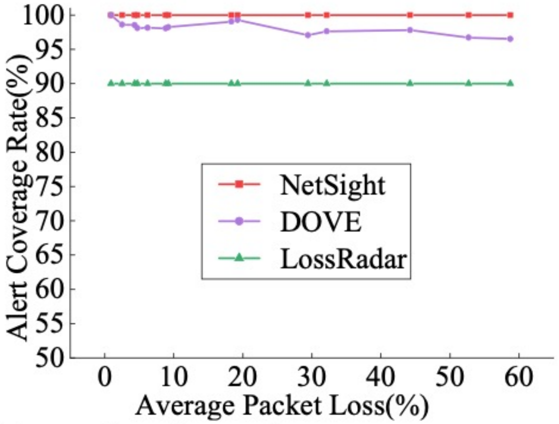
DOVE's overhead: alert, event, telemetry header



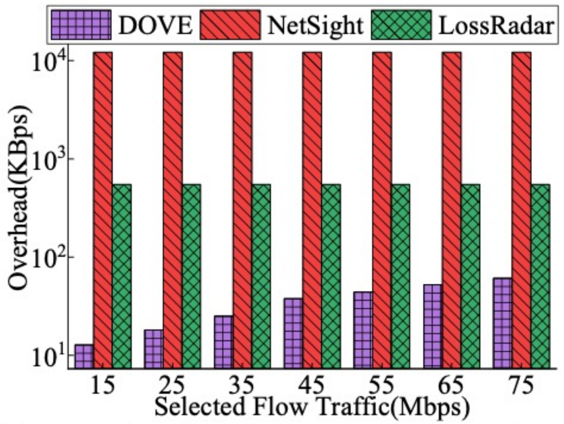
- there is a tradeoff between SLO measure accuracy (epoch length) and overhead
- telemetry header overhead is proportional to the size of selected flows

Evaluation

DOVE's packet loss: accuracy and overhead

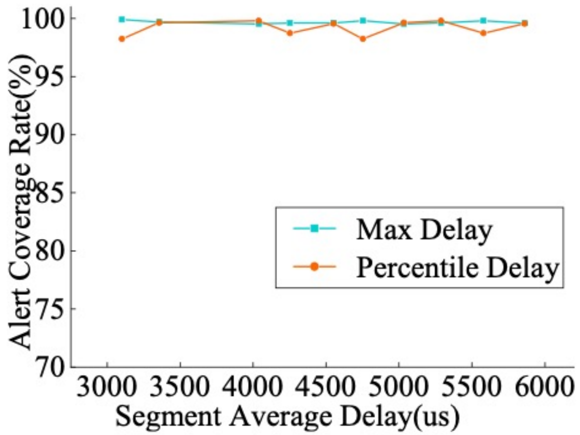


(a) packet loss alert coverage rate with 20 incast flows

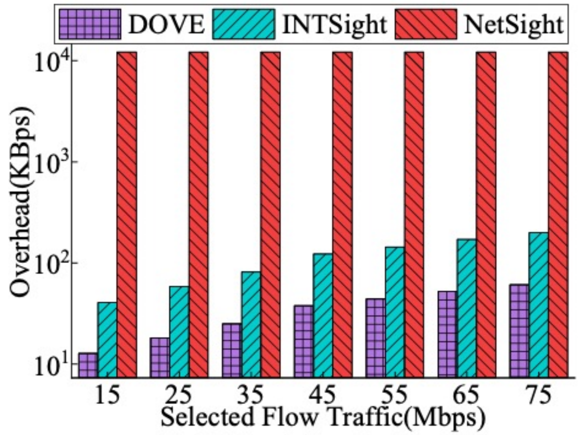


(b) overhead of generating packet loss alerts on 8Gbps link with 1% packet loss

DOVE's delay: accuracy and overhead



(a) max and percentile delay alert coverage



(b) overhead of generating delay alerts on 8Gbps link

- packet loss:
  - good coverage rate (>97%)
  - generates much less traffic overhead compared with NetSight and LossRadar
  - heavy packet loss makes Coloring Algorithm less effective
- delay:
  - generate less traffic overhead than INTSight (simpler telemetry header)

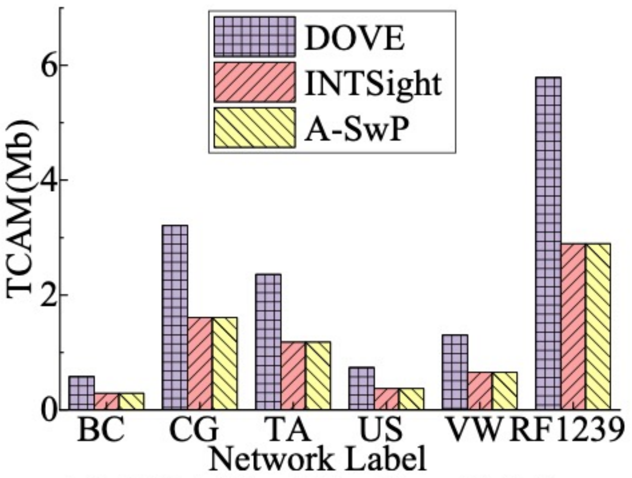
Evaluation

DOVE's resource utilization over large networks

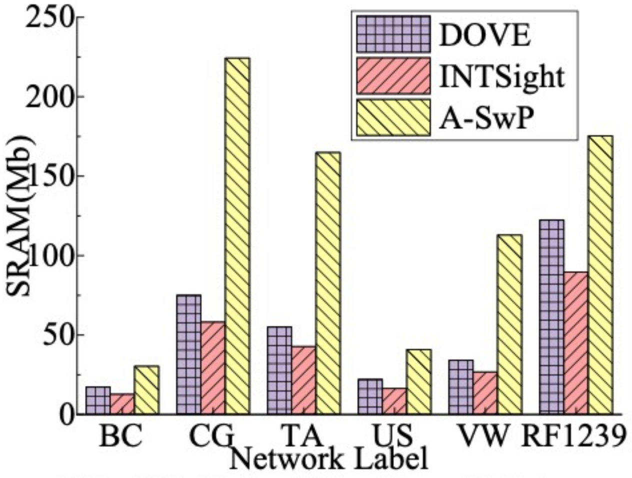
- 512 selected flows and 512 watched flows for each pair of nodes
- DOVE TCAM = 2x INTSight TCAM
  - DOVE monitors two sets of flows as INTSight only monitors one
- DOVE SRAM > INTSight SRAM
  - DOVE requires many registers to store intermediate values
- The required resources can fit into programmable switches such as Tofino.

Metadata of network topologies.

Network	Label	Nodes	Links	Average Path Length
Bell Canada	BC	48	130	5.3
US Signal	US	61	158	6.0
VTLWavenet	VW	92	192	13.1
TATA	TA	145	388	9.9
Cogent	CG	197	490	10.5
RF1239	RF1239	315	1944	4.0



(a) TCAM utilization (Mb)



(b) SRAM utilization (Mb)

Diagnosis-driven SLO Violation DEtection

THANKS

leiyr20[AT]mails[DOT]tsinghua[DOT]edu[DOT]cn