Time-Muxed Parsing in Marking-based Network Telemetry

Alon Riesenberg*, Yonnie Kirzon*, Michael Bunin*,

Elad Galili*, Gidi Navon●, Tal Mizrahi◊*

ACM SYSTOR, Haifa, May 2019

* TECHNION Israel Institute of Technology

● MARVELL®

◊ HUAWEI

Background

What is network telemetry?

Delay | **Packet loss** | **Queue status** | ...

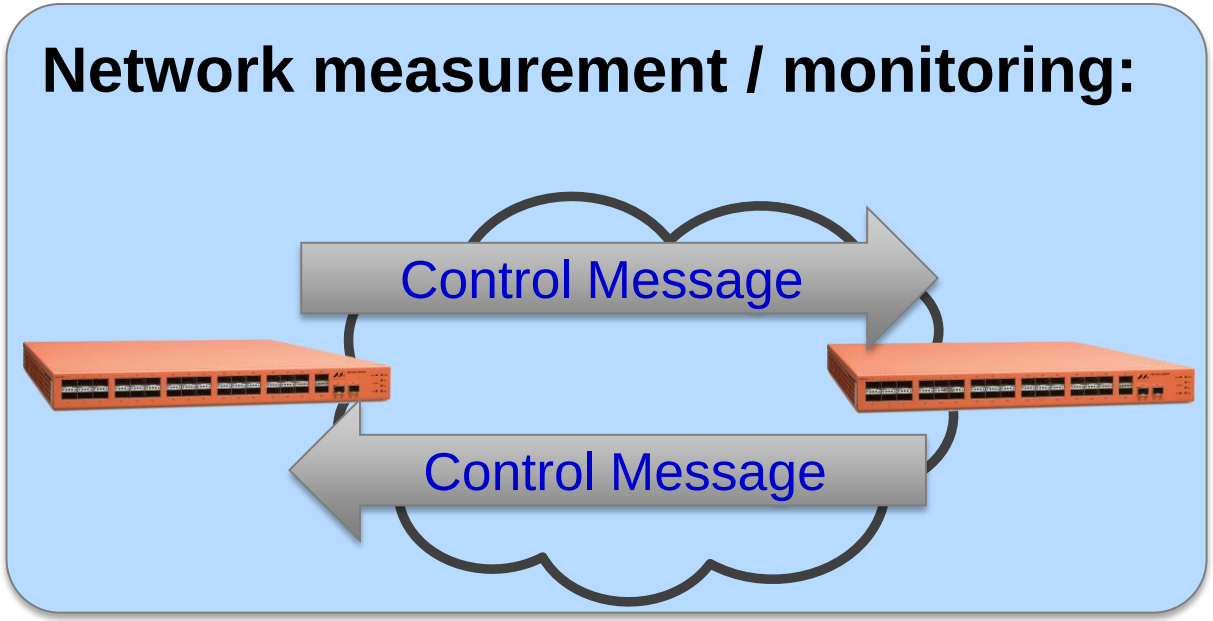
Performance measurement + exporting to a remote location

Why do we need telemetry?

Detection

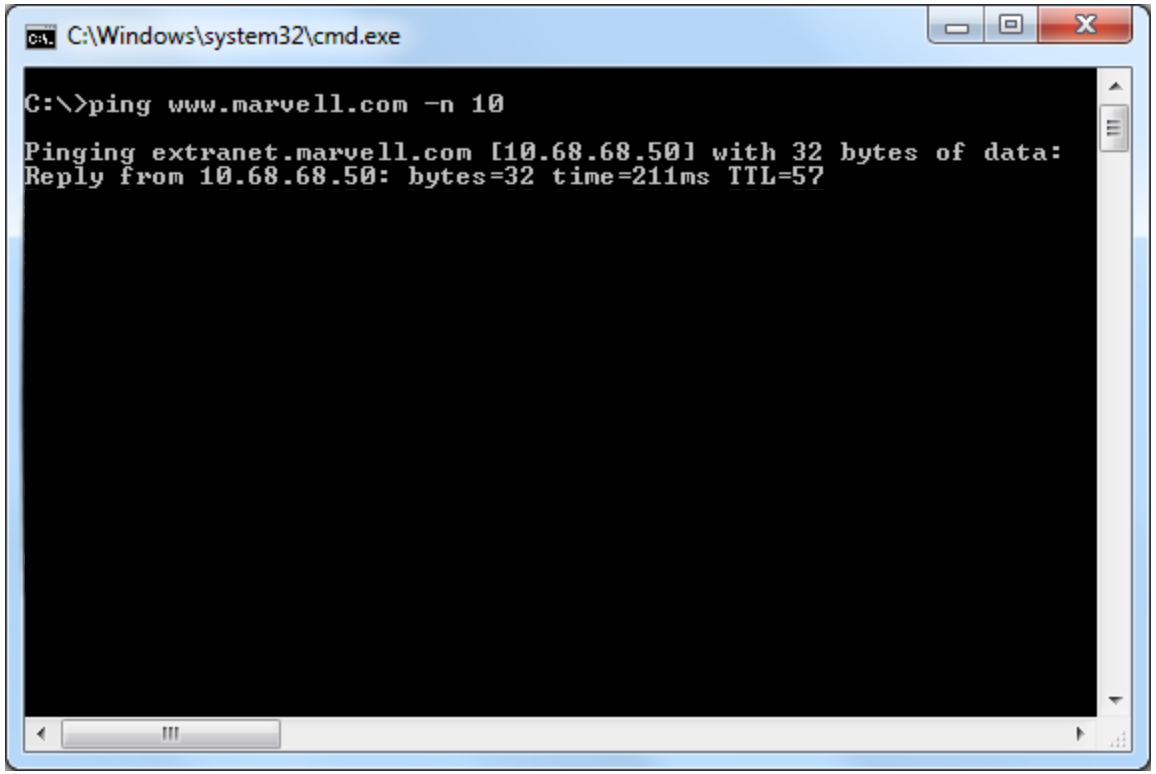
Failures | **Congestion / Bottlenecks** | **'Elephant' flows** | ...

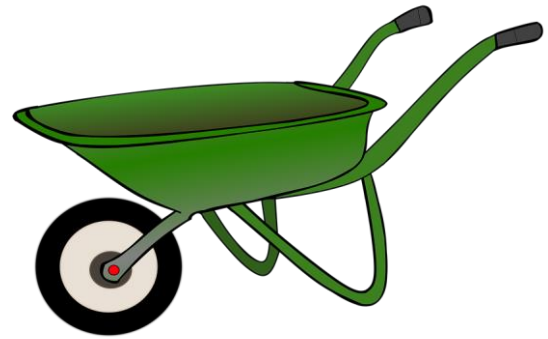
Operations, Administration, Maintenance (OAM)



Network Telemetry

Ping / Traceroute





Old-School Passive Monitoring

Counters

Queue State

Latency

⋮

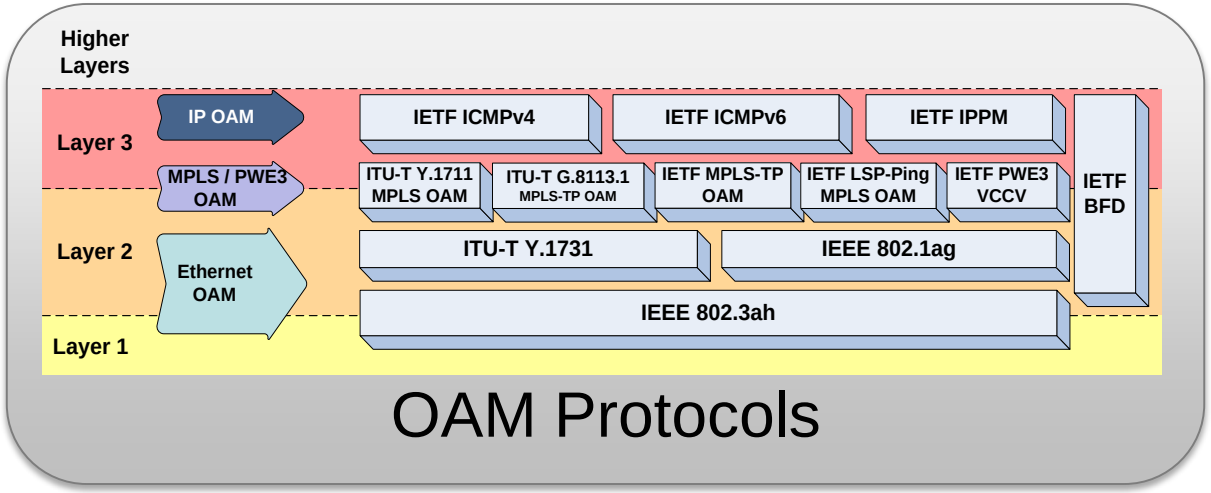
Per port

Per flow

Per queue

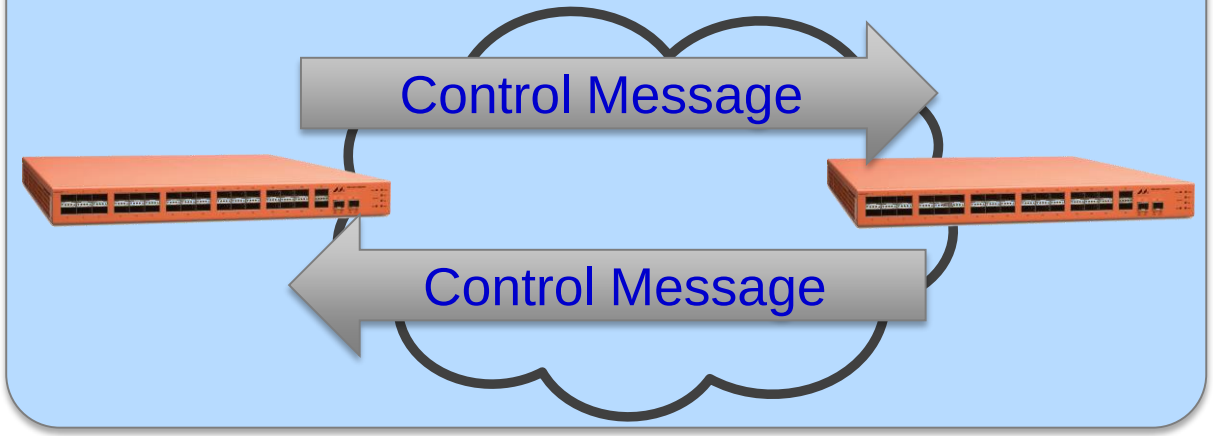
⋮

Carrier Network OAM

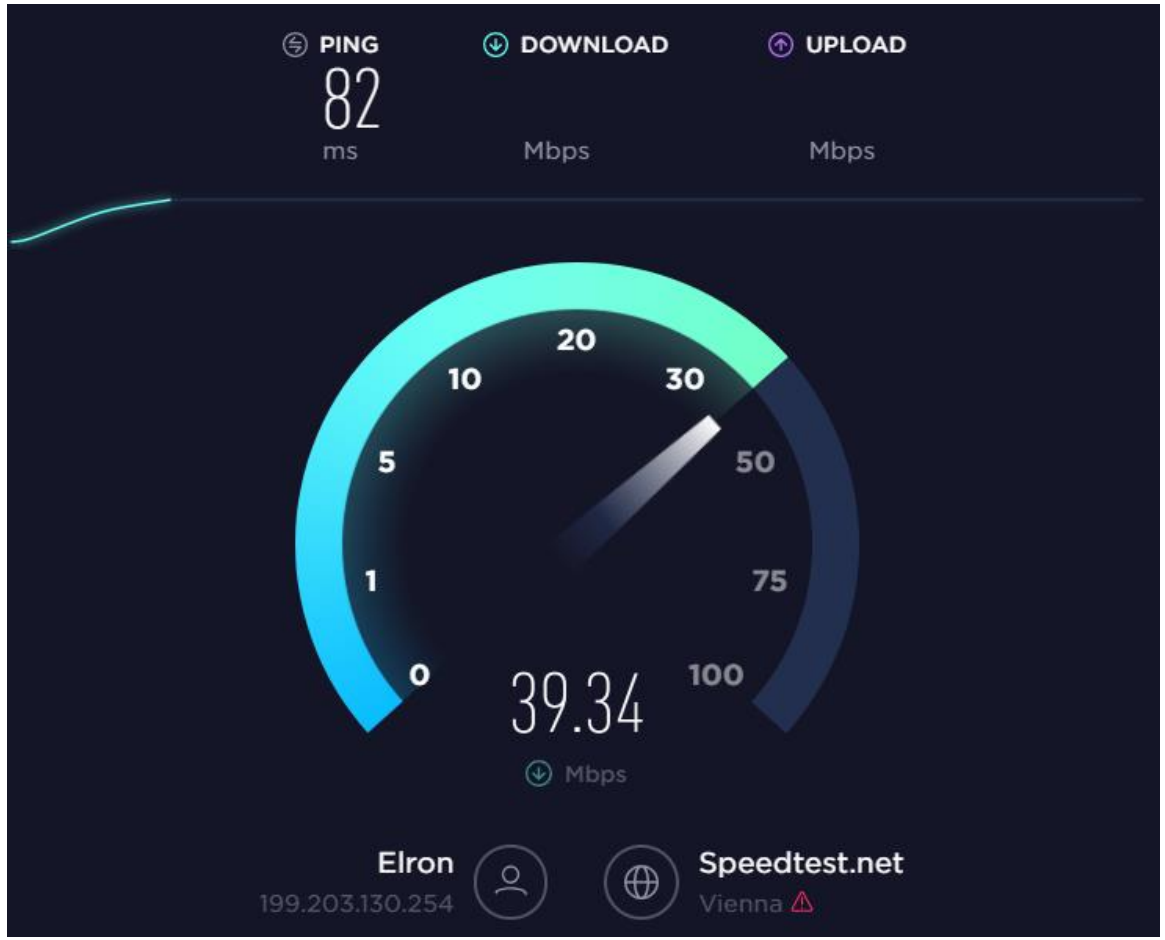


OAM Protocols

Active measurement / monitoring:

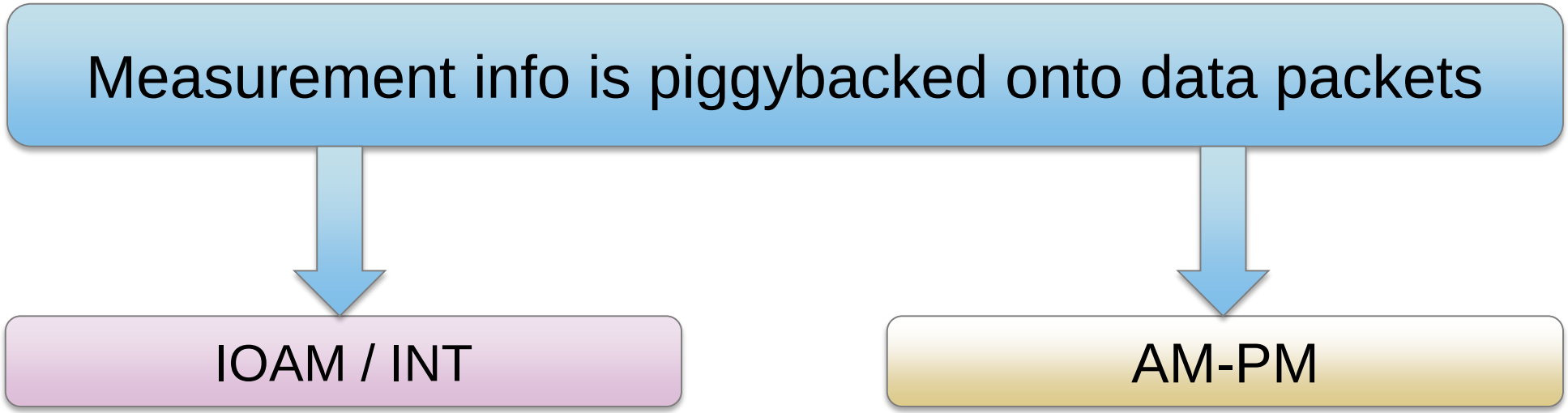


Fate Sharing

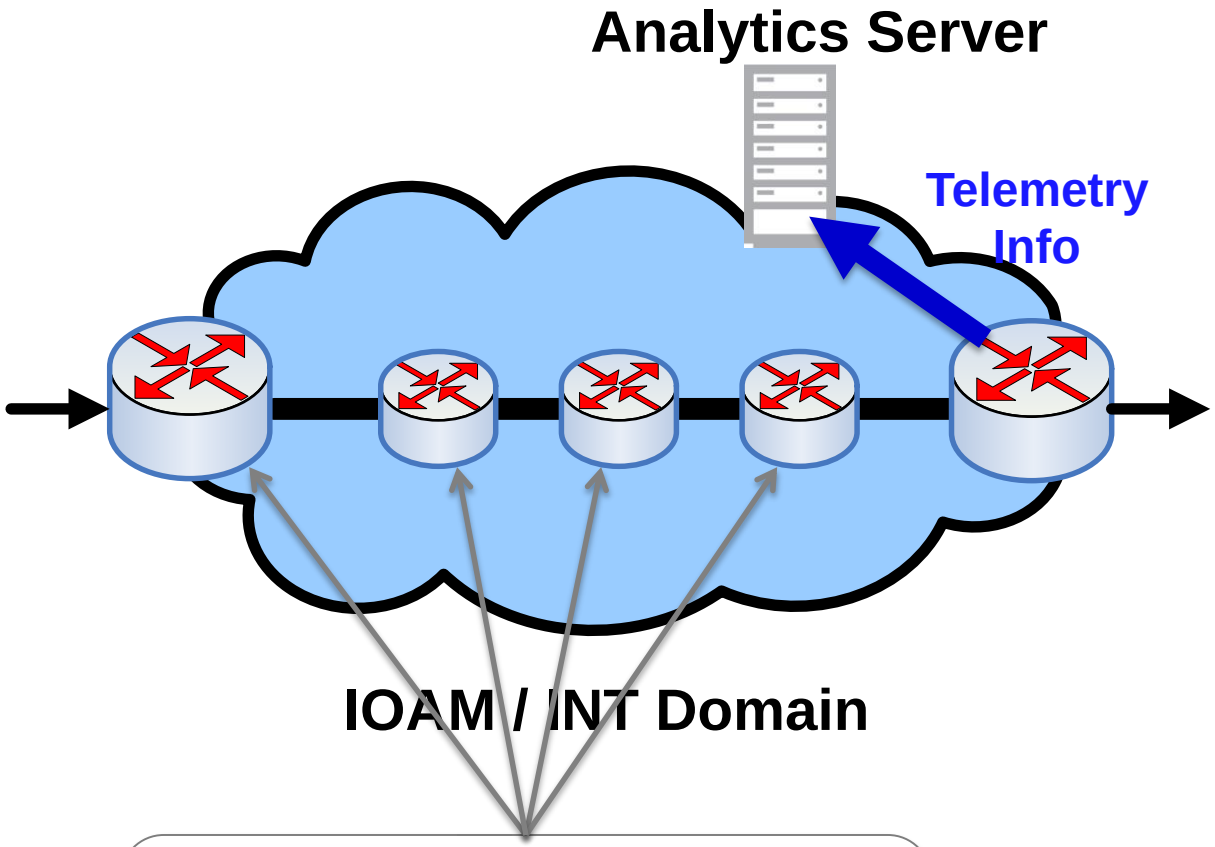


http://www.speedtest.net

Piggybacked Measurement



Piggybacked Metadata – IOAM / INT

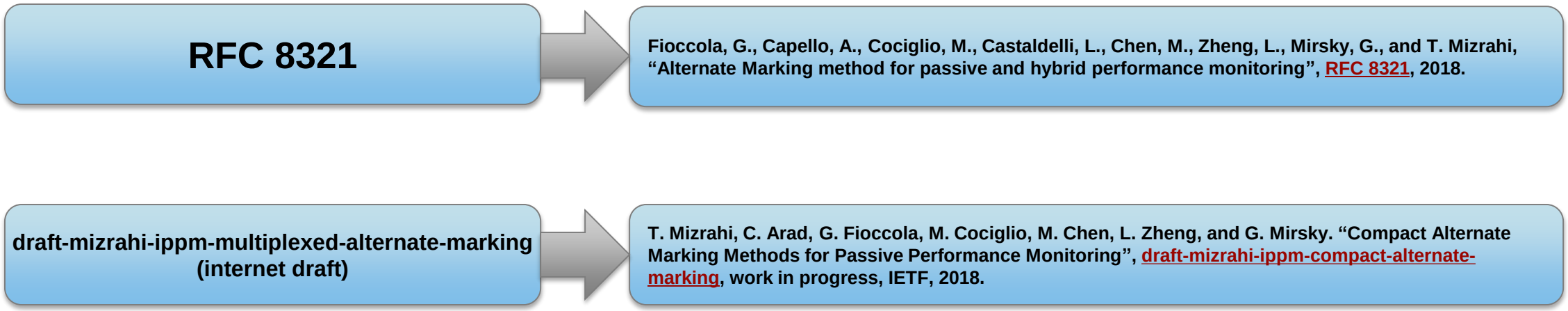


Switches push local metadata into header: delay, queue state, …

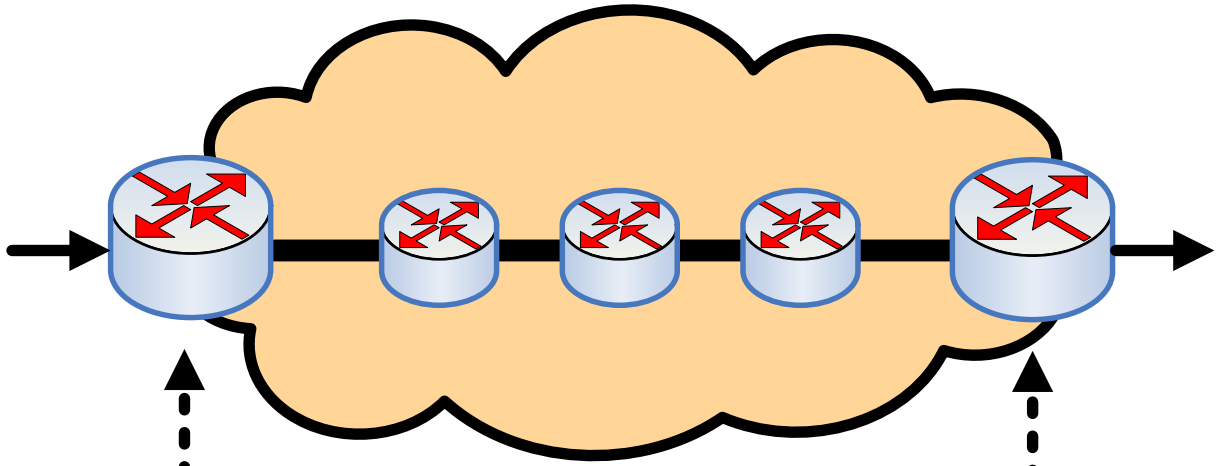
IOAM In situ OAM
INT In-band Network Telemetry

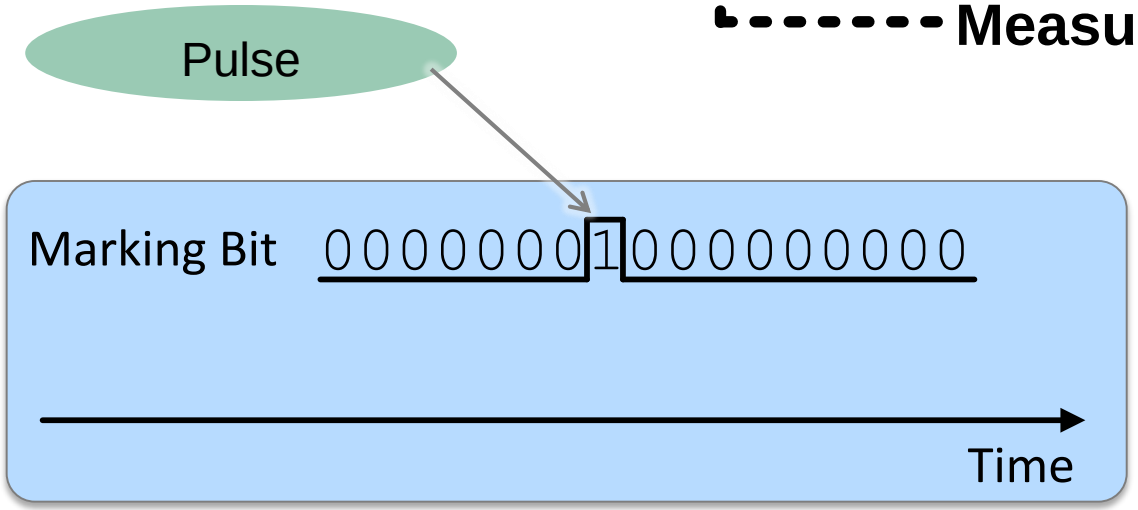
Per-packet metadata ☺
Per-packet overhead ☹

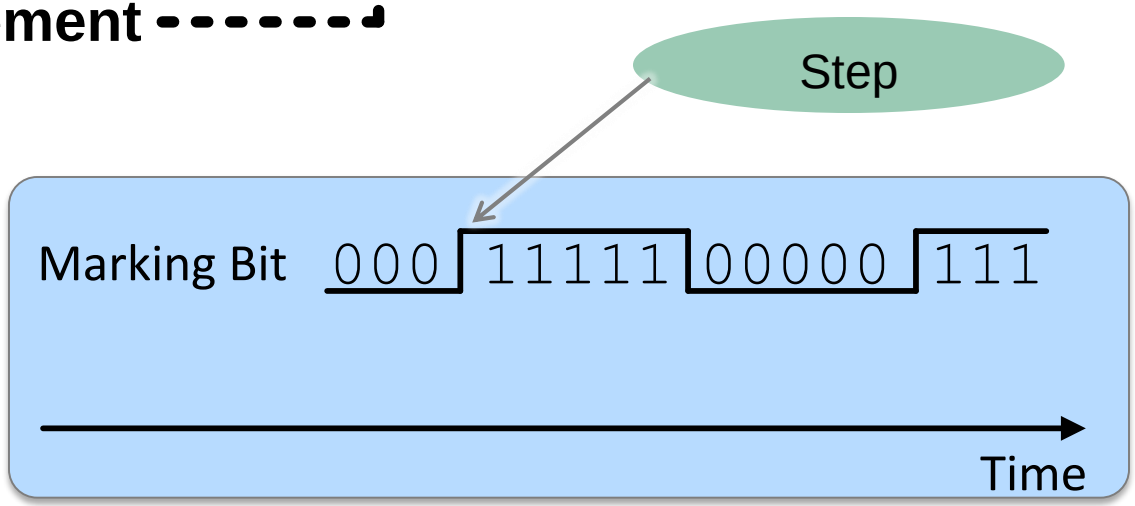
AM-PM: Alternate Marking – Performance Measurement



AM-PM: What Can We Do with ONE Bit Per Packet?







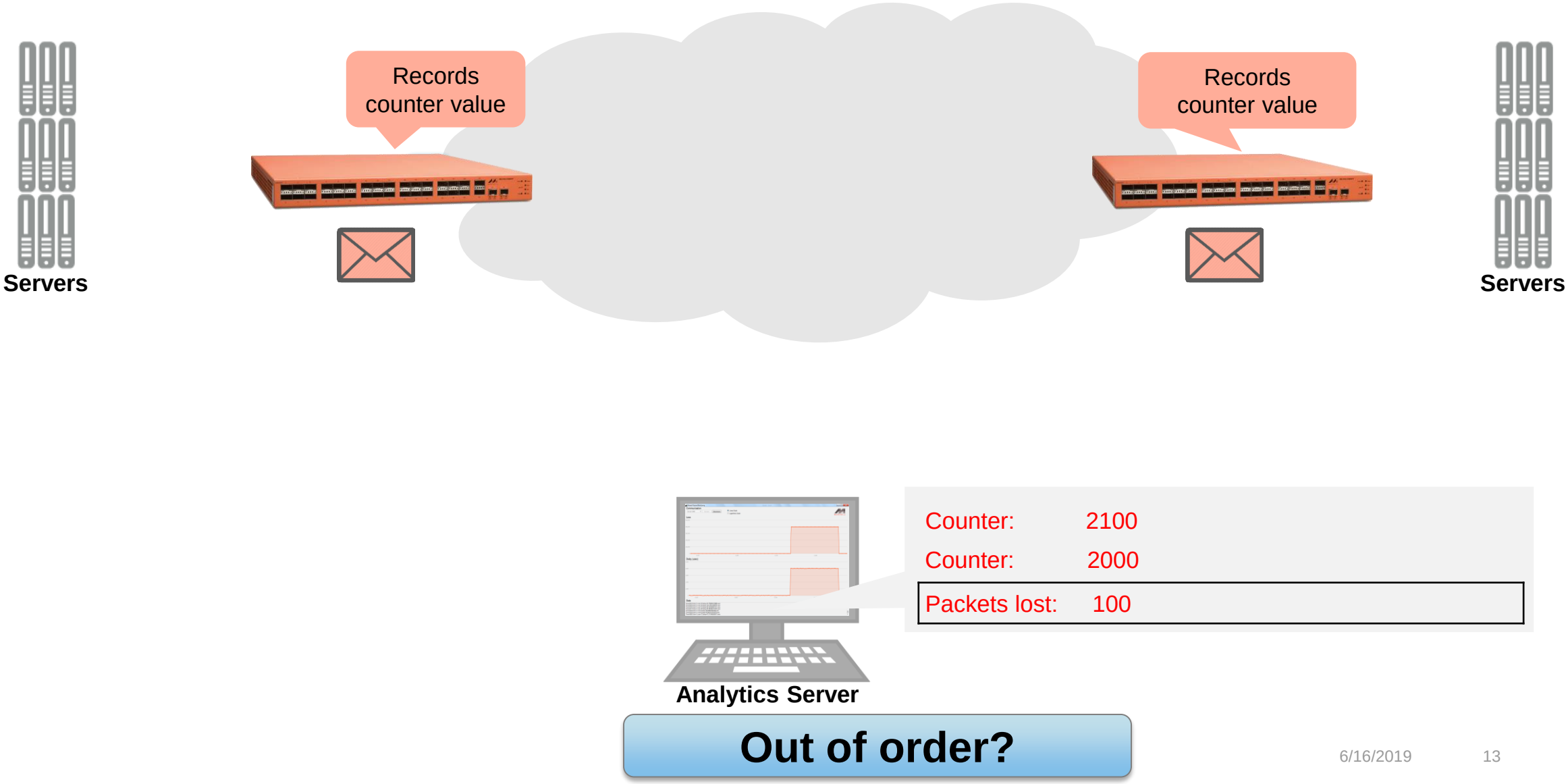
AM-PM: Pulse Marking – Delay Measurement

Time Sent: March 8th, 16:02, 123400789 nsec (UTC)

Time Received: March 8th, 16:02, 123500789 nsec (UTC)

Network Delay: 100 µsec

AM-PM: Pulse Marking – Loss Measurement



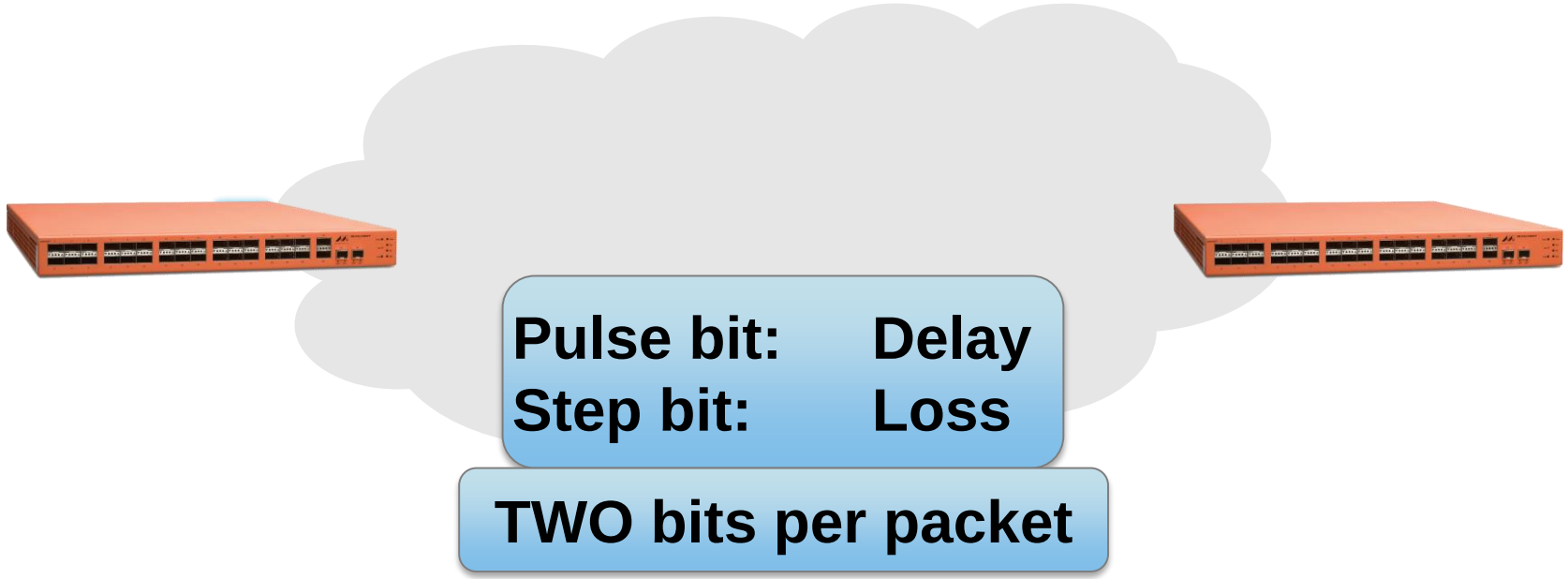
AM-PM: Alternate Marking – Loss Measurement

Consistent counting:

- **Export the counter of each color when it is not in use.**
- **Resilient to reordering.**

AM-PM: Double Marking

Servers



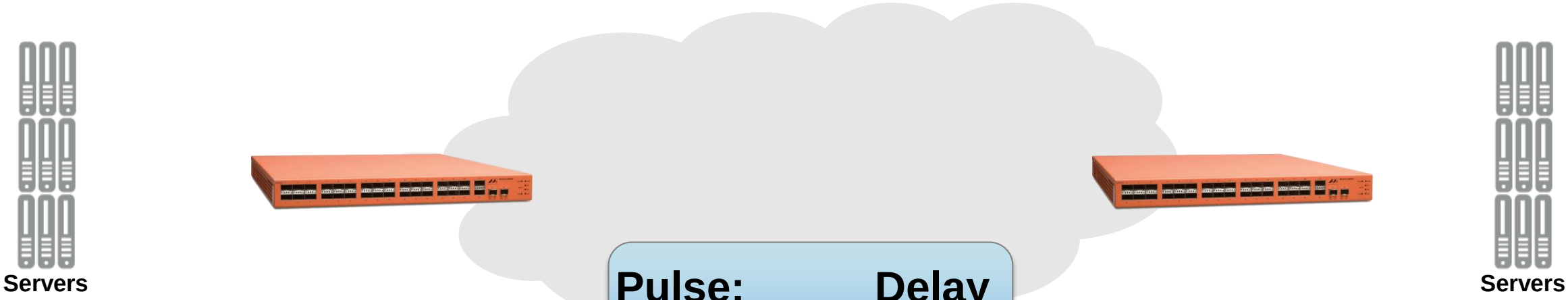
Pulse bit: Delay
Step bit: Loss

TWO bits per packet

Servers

Analytics Server

AM-PM: Multiplexed Marking



Pulse: Delay
Step: Loss

ONE bit per packet

Accurate loss and delay measurement!

Design and Implementation of AM-PM

Match-Action Lookup

TCAM / Exact match / P4

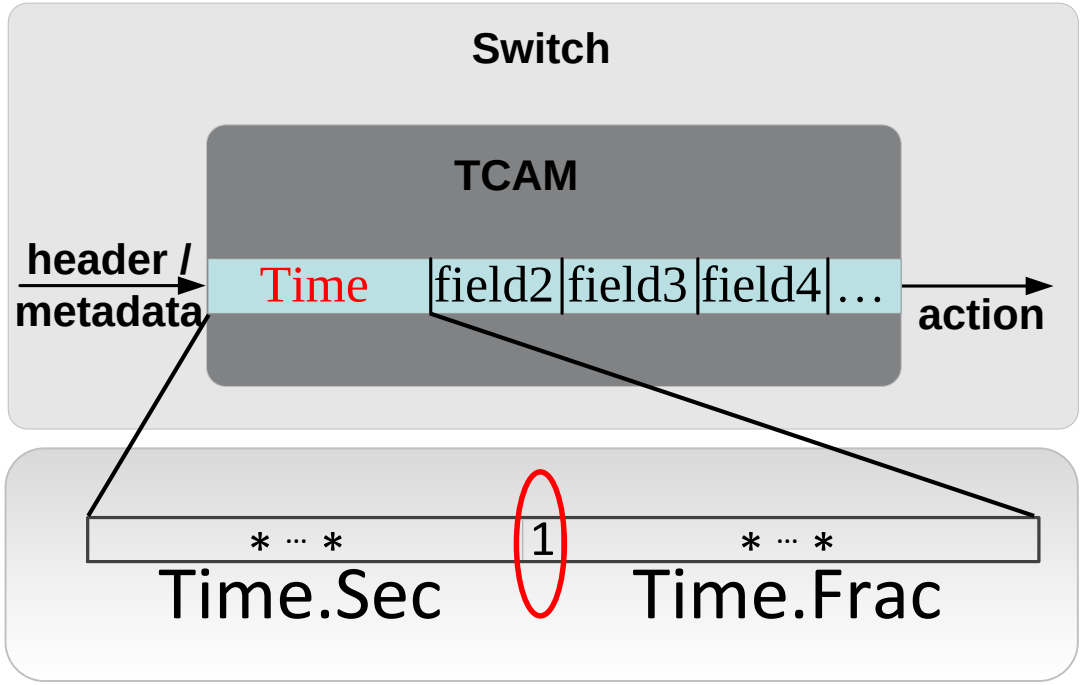
Time-as-a-match

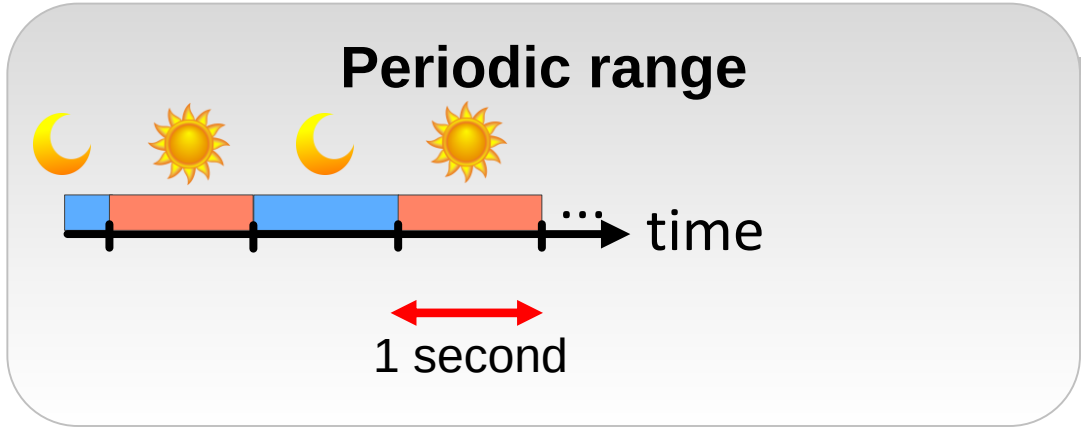
TimeFlip

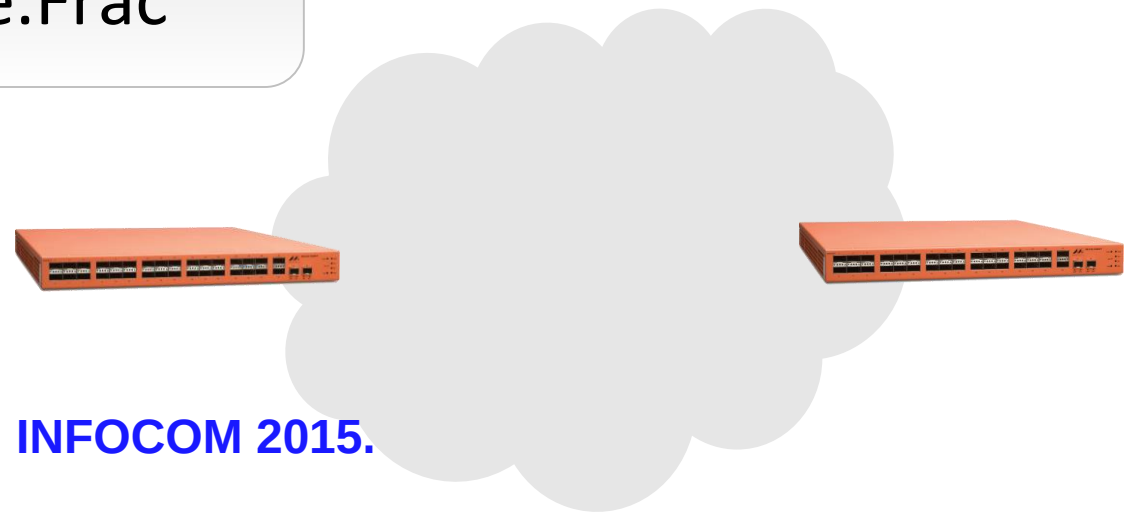
State

Detect first packet (pulse/step)

Time-as-a-match: TimeFlip [MRM]







[MRM] Mizrahi, Rottenstreich, Moses, INFOCOM 2015.

Design and Implementation of AM-PM: Step/Pulse

Match-Action Lookup

TCAM / Exact match / P4

Time-as-a-match

TimeFlip

State

Detect first packet (pulse/step)

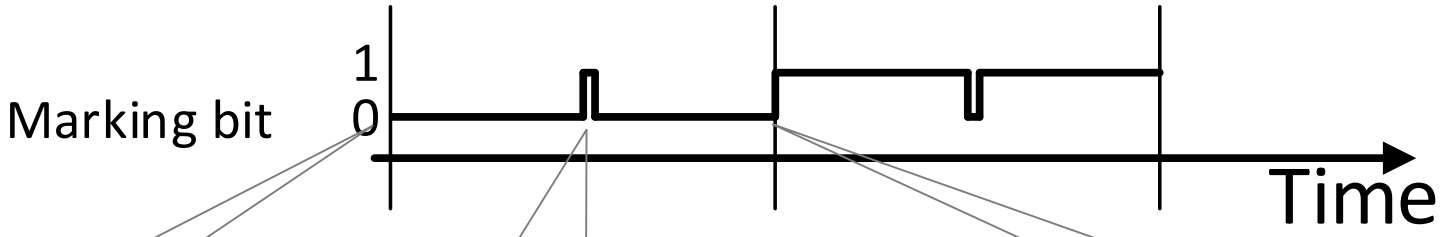
Match	Action
$TimeBit = 0$	$MarkBit = 0, counter0$

Match	Action
$MarkBit = 0$	$counter0$

Match	Action
$TimeBit = 1, Reg = 0$	$MarkBit = 1, Reg = 1, timestamp$
$TimeBit = 1, Reg = 1$	$MarkBit = 0, Reg = 1$
$TimeBit = 0, Reg = 1$	$MarkBit = 1, Reg = 0,$

Match	Action
$MarkBit = 1$	$timestamp$

Multiplexed Marking: a Naïve Implementations



Track the value of the marking bit.

Detect pulse

When the value changes for **one** packet.

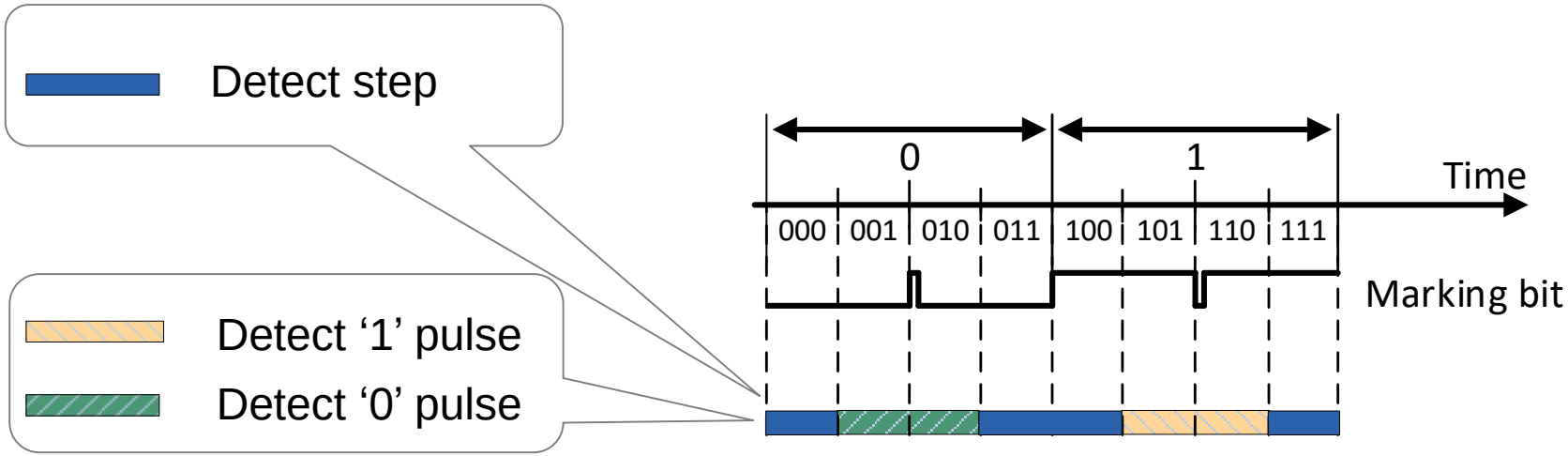
Detect step

When the value changes for more than **one** packet.

Non-trivial to implement using a match-action abstraction.

Our Approach: Time-multiplexed Parsing

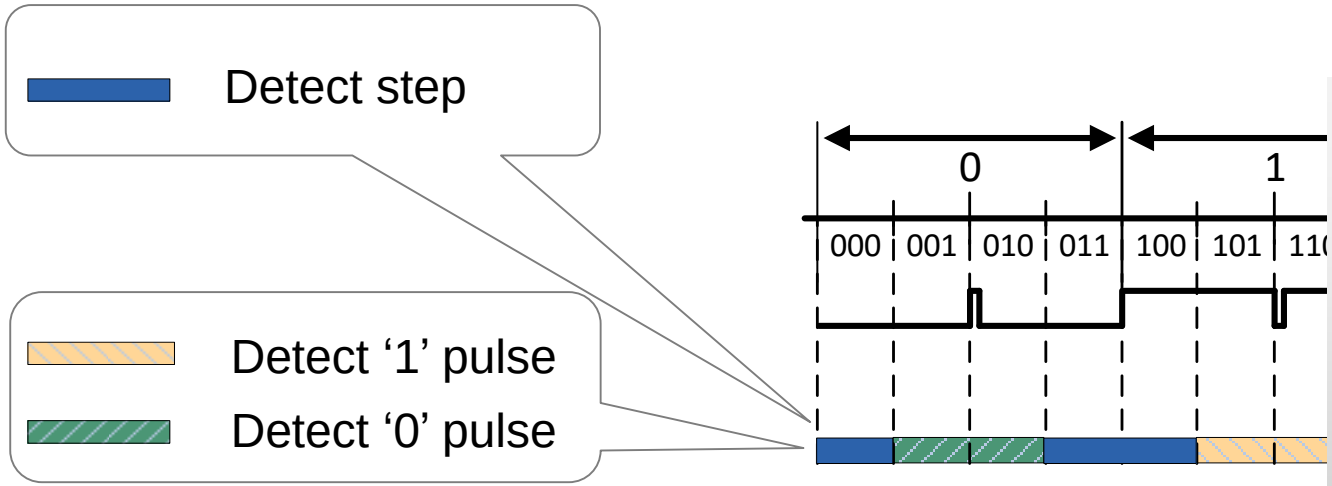
Header field(s) have a different interpretation in each time slot!



- **TimeFlip is used to divide time into time slots.**
- **The marking bit has a different interpretation in each time slot.**
- **Requires rough time synchronization, e.g., ~ 1 second.**

Our Approach: Time-multiplexed Parsing

Header field(s) have a different interpretation in each time slot!

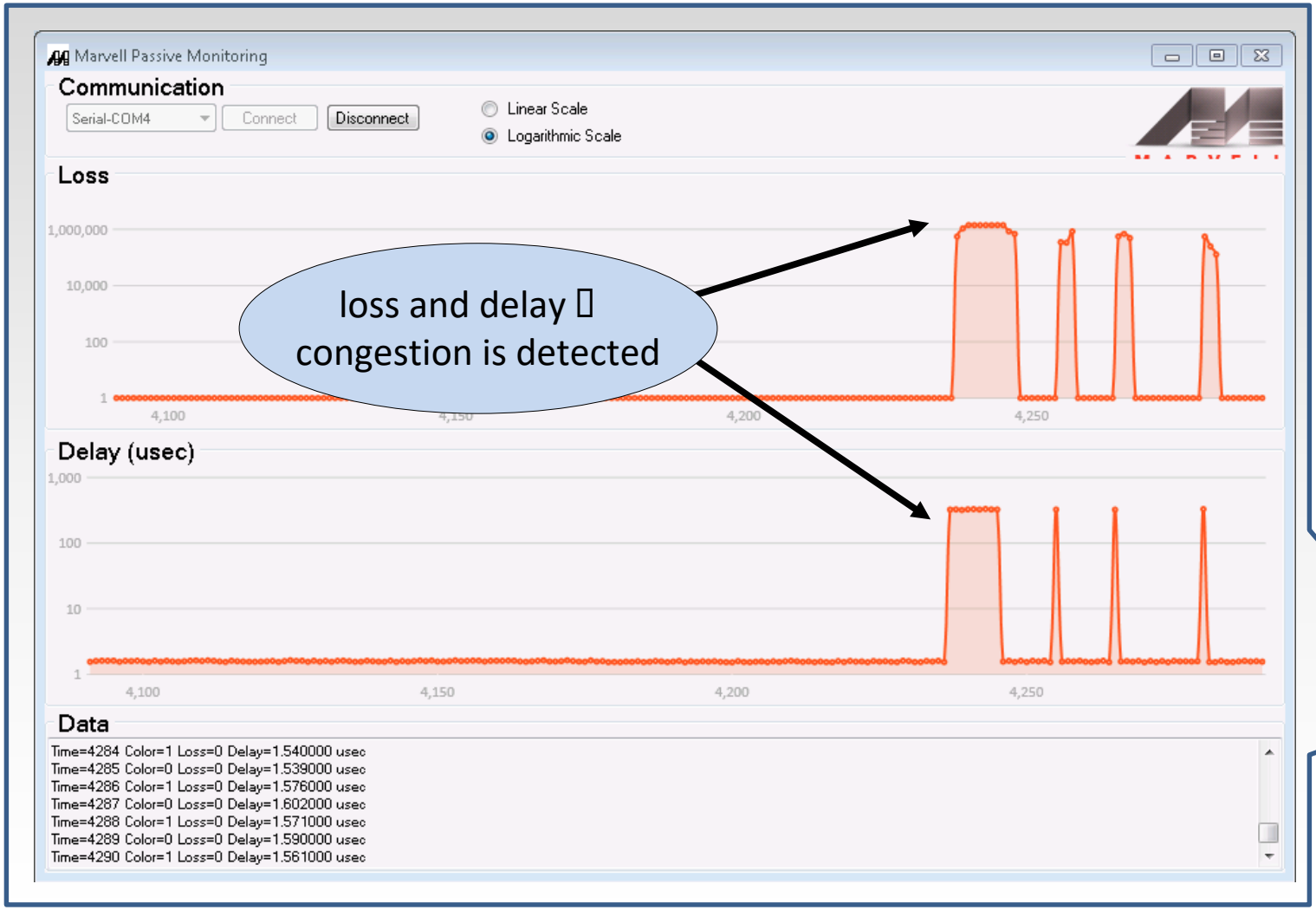


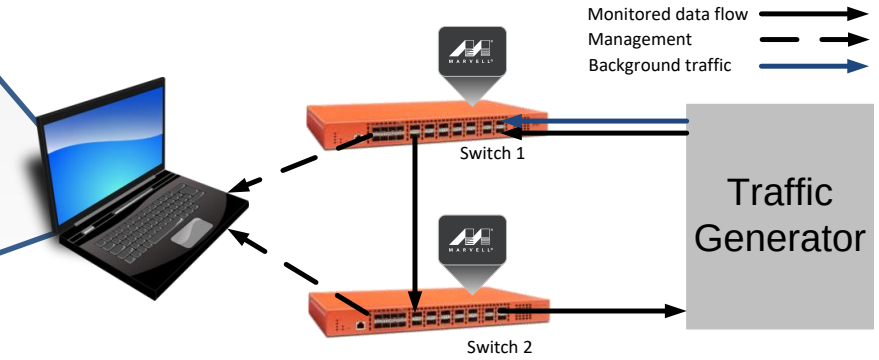
Match	Action
$TimeBits = 010, Rea = 0$	$MarkBit = 1, Rea = 1$

Match	Action
$TimeBits = 001$, $MarkBit = 1$	$counter0$, $timestamp$
$TimeBits = 010$, $MarkBit = 1$	$counter0$, $timestamp$
$TimeBits = 101$, $MarkBit = 0$	$counter1$, $timestamp$
$TimeBits = 110$, $MarkBit = 0$	$counter1$, $timestamp$
$TimeBits = ***$, $MarkBit = 0$	$counter0$
$TimeBits = ***$, $MarkBit = 1$	$counter1$

- **TimeFlip is used to divide time into time**
- **The marking bit has a different interpretation in each time slot.**
- **Requires rough time synchronization, e.g., ~ 1 second.**

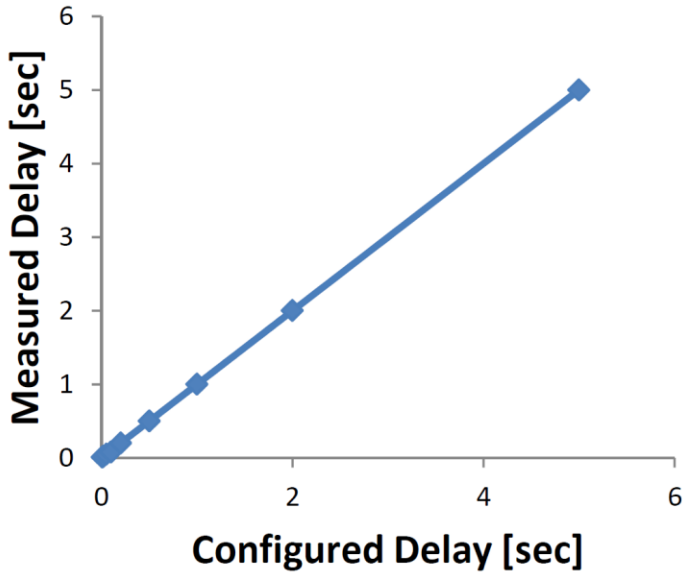
AM-PM Evaluation using Marvell Prestera Switches

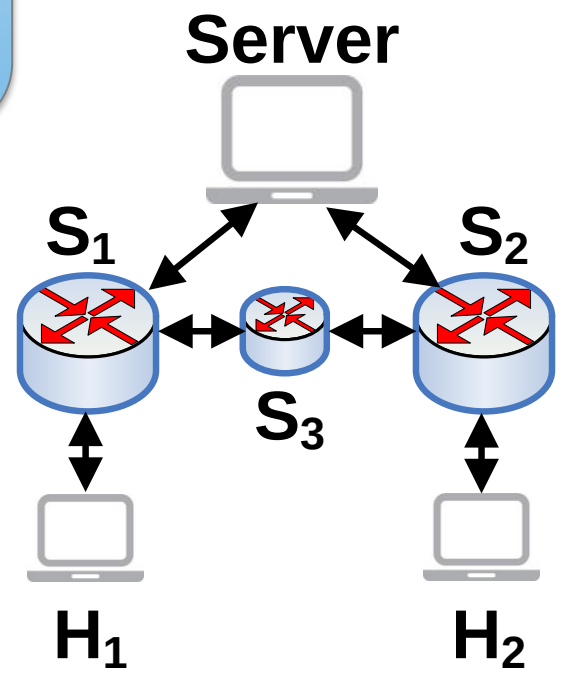




Software Implementation using P4

- Implemented in P4.
  - Time-of-day match field.
  - AM-PM in P4.
- Tested in Mininet.
- Open source code.



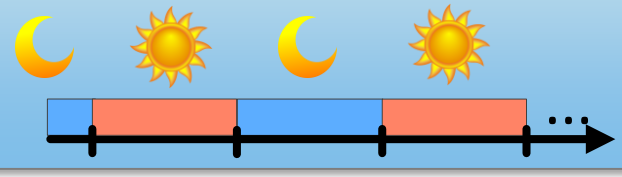


AM-PM: Where is it going?

Network telemetry

Low overhead

AM-PM



Ongoing AM-PM work in the IETF:

QUIC

MPLS

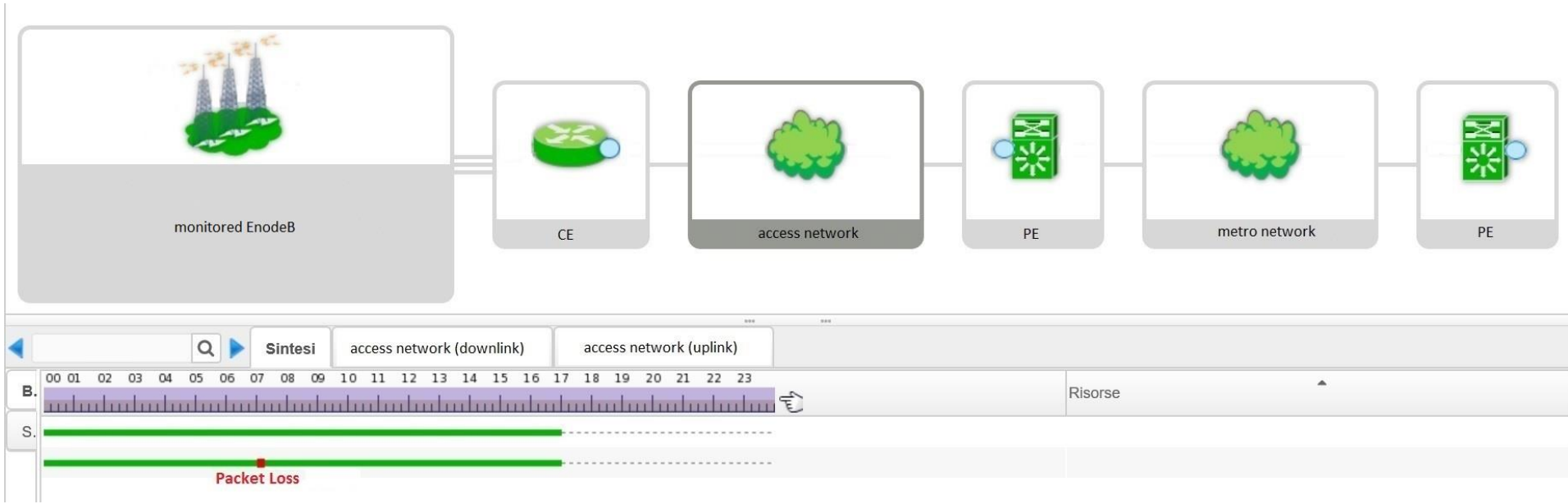
NSH

BIER

Geneve

AM-PM is under discussion in 6 working groups in the IETF…

Large Scale Deployment in Telecom Italia



- **Mobile backhaul network ~ 1000 eNodeBs.**
- **AM-PM one bit (step-based) loss measurement.**
- **Uses unused bit in DSCP.**
- **Off-the-shelf network equipment.**

Summary

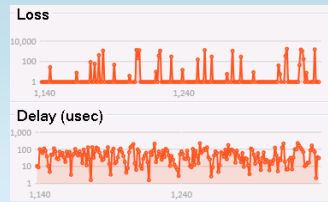
Design and implementation of AM-PM

Hardware-based implementation using a Marvell switch.

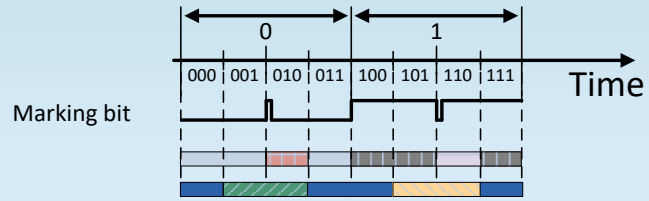
Software-based implementation in P4 – open source.

```
table look_for_flag {
    reads {
        intrinsic_metadata.time_of_day : ternary;
            ipv4.flag_a : exact;
    }
    actions {
            _look_for_flag;
        _drop;
    }
    size: 256;
}
```

Experimental results



Novel time-multiplexed parsing




12
1
2
3
4
5
6
7
8
9
10
11
13
14
15
16
17
18
19
20
21
22
23
24
Thanks!

References

[1] Fioccola, G., Capello, A., Cociglio, M., Castaldelli, L., Chen, M., Zheng, L., Mirsky, G., and T. Mizrahi, “Alternate Marking method for passive and hybrid performance monitoring”, RFC 8321, 2018.

[2] Mizrahi, T., Arad, C., Fioccola, G., Cociglio, M., Chen, M., Zheng, L., and G. Mirsky, “Compact Alternate Marking Methods for Passive and Hybrid Performance Monitoring”, draft-mizrahi-ippm-compact-alternate-marking, work in progress, IETF, 2019.

[3] Brockners, F., Bhandari, S., Pignataro, C., Gredler, H., Leddy, J., Youell, S., Mizrahi, T., Mozes, D., Lapukhov, P., Chang, R. and D. Bernier, J. Lemon, "Data Fields for In-situ OAM", draft-ietf-ippm-ioam-data, work in progress, 2019.

[4] C. Kim et al., “In-band network telemetry (INT)”, P4 consortium, 2015.

[5] Mizrahi, T., Vovnoboy, V., Nisim, M., G. Navon, and A. Soffer, “Network Telemetry Solutions for Data Center and Enterprise Networks”, Marvell white paper, 2018.

[6] Mizrahi, T., Rottenstreich, O. and Y. Moses, “TimeFlip: Scheduling Network Updates with Timestamp-based TCAM Ranges”, IEEE INFOCOM, 2015.

[7] Mizrahi, T., Navon, G., Fioccola, G., Cociglio, M., Chen, M., and G. Mirsky, “AM-PM: Efficient Network Telemetry using Alternate Marking”, IEEE Network, 2019.

[8] Riesenberg, A., Kirzon, Y., Bunin, M., Galili, E., Navon, G., and T. Mizrahi, “Time-Multiplexed Parsing in Marking-based Network Telemetry”, ACM SYSTOR, 2019.

[9] P4 AM-PM, https://github.com/AlternateMarkingP4/FlaseClase, 2018.