Time-Muxed Parsing in Marking-based Network Telemetry

Alon Riesenbergs*, Yonnie Kirzon*, Michael Bunin*,
Elad Galilis*, Gidi Navons*, Tal Mizrahi†*

ACM SYSTOR, Haifa, May 2019
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
- ‘Elephant’ flows
Operations, Administration, Maintenance (OAM)

Network measurement / monitoring:

Control Message

Control Message

Network Telemetry
Ping / Traceroute

C:\\Windows\system32\cmd.exe

C:\>ping www.marvell.com -n 10
Pinging extranet.marvell.com [10.68.60.58] with 32 bytes of data:
Reply from 10.68.60.58: bytes=32 time=21ms TTL=57

0.0.0.0

6/16/2019
Old-School Passive Monitoring

Counters Per port
Queue State Per flow
Latency Per queue
Carrier Network OAM

OAM Protocols

Active measurement / monitoring:

Control Message

Control Message
Fate Sharing

http://www.speedtest.net
Piggybacked Measurement

Measurement info is piggybacked onto data packets

IOAM / INT

AM-PM
Piggybacked Metadata – IOAM / INT

Switches push local metadata into header: delay, queue state, …
AM-PM: Alternate Marking – Performance Measurement

RFC 8321


draft-mizrahi-ippm-multiplexed-alternate-marking (internet draft)

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

Marking Bit 00000001000000000

Marking Bit 0001111000000111

Pulse

Step

Measurement

Time
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

Counter: 2100
Counter: 2000
Packets lost: 100

Out of order?
AM-PM: **Alternate Marking** – Loss Measurement

**Analytics Server**

- Counts number of packets sent
- Counts number of packets received

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

**Packets**
- Sent: 10,000
- Lost: 500
- Received: 9,500

**Dates:** 6/16/2019 14
AM-PM: Double Marking

Pulse bit: Delay
Step bit: Loss
TWO bits per packet
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

State
Detect first packet (pulse/step)

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

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)

---

<table>
<thead>
<tr>
<th>Match</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>TimeBit = 0</td>
<td>MarkBit = 0, counter0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Match</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>MarkBit = 0</td>
<td>counter0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Match</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>TimeBit = 1, Reg = 0</td>
<td>MarkBit = 1, Reg = 1, timestamp</td>
</tr>
<tr>
<td>TimeBit = 1, Reg = 1</td>
<td>MarkBit = 0, Reg = 1</td>
</tr>
<tr>
<td>TimeBit = 0, Reg = 1</td>
<td>MarkBit = 1, Reg = 0, timestamp</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Match</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>MarkBit = 1</td>
<td>timestamp</td>
</tr>
</tbody>
</table>

---
Multiplexed Marking: a Naïve Implementations

Marking bit

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!

- **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.
AM-PM Evaluation using Marvell Prestera Switches

Loss and delay in congestion is detected

Traffic Generator
Management
Monitored data flow
Background traffic

Switch 1
Switch 2

Traffic Generator
Software Implementation using P4

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

![Graph showing measured delay versus configured delay](image)

![Diagram showing network topology](image)
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.

Experimental results

Novel time-multiplexed parsing
Thanks!
References


