Advancing Metrics on the Standards Track:
RFC 2680 (1-way Loss)
Test Plan and Results

draft-ietf-ippm-testplan-rfc2680-02
Len Ciavattone, Rüdiger Geib,
Al Morton, Matthias Wieser
March 2013
Outline

- Implement the Definition-centric metric advancement described in RFC 6576
- Test Plan Overview
  - Test Set-up and Specific Tests
- Test Results
- Summary and implications on the text of the revised RFC2680
Definition-Centric Process

( Start )
\  /  
\-+-' Implementations
  /   |
+/-----+ 1   `-+-'
   /   `---------.-------
RFC / Check for ', was RFC ' . YES
   / |Equivalence..... clause x ---------+
   |/     |under       |' . clear? ',      |
Metric \.....| 2 ....relevant |'------' ---------+
Metric | identical | No | Report |
Metric | network | +-----+ results+|
... | conditions | Modify | Advance |
|   |                     | Spec   +-----+RFC |
|    |                     |         | request? |
+---------\| n |.'+---------+

+---------
Test Configuration

VLAN 200
VLAN 100

Lo0=193.159.144.8
Internet
Lo0=192.168.50.211
NAT=12.3.167.16

MS-1
VLAN 300
192.168.50.201
10.200.0.1

MS-2
VLAN 300
192.168.50.202
10.200.0.2

MS-3
VLAN 400
192.168.50.203
10.200.0.3

MS-4
VLAN 400
192.168.50.204
10.200.0.4

L2TPv3 Tunnel Head
10.200.0.1
10.200.0.2
10.200.0.3
10.200.0.4

VLAN 100
VLAN 200

Sender for Perfas 1 and Perfas 2 flows

Sender for Perfas 3 and Perfas 4 flows

Net Mgt LAN

VLAN 300

MS-1

L2TPv3 Tunnel Head

VLAN 400

Net Mgt LAN

Sender for S1 + S2 flows

Sender for SA + SB flows

Receiver for SA + SB flows

Sender for Perfas 3 and Perfas 4 flows

Perfas+

NetProbe

Sender for S1 + S2 flows

Sender for SA + SB flows
Overview of Testing

- 32 different experiments conducted from March 9 through May 2, 2011.
- Varied Packet size, Active sampling distribution, test duration, and other parameters (Type-P)
- Added Network Emulator “netem” and varied fixed and variable delay distributions
  - Inserted loss in a limited number of experiments.
Results Summary (details in memo)

- Loss Counts – **Pass ADK** (adj for ties), 3 conditions
- Calibration – completed for both implementations
- Loss Threshold – available in post-processing for both implementations (used results in RFC2679 plan)
  - Suggest revised text to allow this in RFC
- Loss with Reordering
  - Netem independent delay 2 sec +/- 1 sec
  - Loss Counts **Pass ADK** as before.
- Poisson Distribution AD GoF, multiple sample sizes
  - Both NetProbe and Perfas pass in both sample sizes
- Delay Stats – There’s only one:
  - Both Implementations report (as loss ratio)
  - Type-P-One-way-Loss-Average <= revise to -Ratio
Revisions in 02 (01 pub in 2013)

- Mostly from IESG feedback on 2679 test plan
- Add “This is supporting info, not the text of 2680bis” paragraph (the revised text exists!)
- Added References for NetProbe and Perfas+
  - Perfas+ ref in German
- New section describing all conclusions from testing
- The need to address 2680 Errata now included
Summary

- Two Implementations: NetProbe and Perfas+
- Test Plan for Key clauses of RFC 2680
  - the basis of Advance RFC Request
  - Criteria for Equivalence Threshold & correction factors
- Experiments complete, key clauses of RFC2680 evaluated
  - Two revisions to the RFC suggested from this study
References


Overview of Testing (sample)

<table>
<thead>
<tr>
<th>Date</th>
<th>Samp</th>
<th>Interval</th>
<th>Duration</th>
<th>Notes</th>
<th>ADK same</th>
<th>ADK cross</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mar 23</td>
<td>Poisson</td>
<td>1s</td>
<td>300s</td>
<td>Netem 10% Loss</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mar 24</td>
<td>Periodic</td>
<td>1s</td>
<td>300s</td>
<td>Netem 100ms +/- 50ms delay</td>
<td></td>
<td>Pass</td>
</tr>
<tr>
<td>Mar 24</td>
<td>Periodic</td>
<td>1s</td>
<td>300s</td>
<td>Netem 10% Loss</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mar 28</td>
<td>Periodic</td>
<td>1s</td>
<td>300s</td>
<td>Netem 100ms</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mar 29</td>
<td>Periodic (rand st.)</td>
<td>1s</td>
<td>300s</td>
<td>Netem 100ms +/- 50ms delay, 64 Byte</td>
<td>NP s12AB Per p1234</td>
<td>Pass combined</td>
</tr>
<tr>
<td>Apr 6</td>
<td>Periodic (rand st.)</td>
<td>1s</td>
<td>300s</td>
<td>Netem 100ms +/- 50ms delay, 340 Byte</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Apr 7</td>
<td>Periodic (rand st.)</td>
<td>1s</td>
<td>1200s</td>
<td>Netem 10% Loss</td>
<td></td>
<td>Pass</td>
</tr>
<tr>
<td>Apr 12</td>
<td>Periodic (rand st.)</td>
<td>1s</td>
<td>300s</td>
<td>Netem 100ms, 500 Byte and 64 Byte comparison</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Criteria for the Equivalence Threshold and Correction Factors

- **Purpose**: Evaluate Specification Clarity (using results from implementations)

- **For ADK comparison**: cross-implementations
  - 0.95 confidence factor at 1ms resolution, or
  - The smallest confidence factor & res. of *same* Implementation

- **For Anderson-Darling Goodness-of-Fit (ADGoF) comparisons**:
  - the required level of significance for Goodness-of-Fit (GoF) SHALL be 0.05 or 5%, as specified in Section 11.4 of [RFC2330]
  - This is equivalent to a 95% confidence factor
Tests in the Plan

6. Tests to evaluate RFC 2680 Specifications
   6.1. One-way Loss, ADK Sample Comparison
       - 64 and 340 Byte sizes
       - Periodic and Poisson Sampling
   6.2. One-way Loss, Delay threshold
   6.3. One-way Loss with Out-of-Order Arrival
   6.4. Poisson Sending Process Evaluation
   6.5. Implementation of Statistics for One-way Delay – Should be Loss
ADK for Loss Counts with 10% netem loss – Cross-Implementations

Null Hypothesis:
All samples within a data set come from a common distribution. The common distribution may change between data sets.

<table>
<thead>
<tr>
<th></th>
<th>ti.obs</th>
<th>P-value*</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>340B 1s Periodic</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>not adj. for ties</td>
<td>0.52043</td>
<td>0.20604</td>
<td></td>
<td></td>
</tr>
<tr>
<td>adj. for ties</td>
<td>0.62679</td>
<td>0.18607</td>
<td></td>
<td></td>
</tr>
<tr>
<td>64B 1s Periodic</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>not adj. for ties</td>
<td>0.76921</td>
<td>0.16200</td>
<td></td>
<td></td>
</tr>
<tr>
<td>adj. for ties</td>
<td>0.90935</td>
<td>0.14113</td>
<td></td>
<td></td>
</tr>
<tr>
<td>64B 1s Poisson**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>not adj. for ties</td>
<td>2.15099</td>
<td>0.04145</td>
<td></td>
<td></td>
</tr>
<tr>
<td>adj. for ties</td>
<td>1.93129</td>
<td>0.05125</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Green = passed, Red = failed
* Some sample sizes < 5, P-value may not be very accurate  
** Streams made two-passes through a netem emulator
Other Results (details in the memo)

- Calibration – completed for both implementations
- Loss Threshold – available in post-processing for both implementations (used results in RFC2679 plan)
  - Suggest revised text to allow this in RFC
- Loss with Reordering
  - Netem independent delay 2 sec +/- 1 sec
  - Loss Counts Pass ADK as before.
- Poisson Distribution AD GoF, multiple sample sizes
  - Both NetProbe and Perfas pass in both sample sizes
- Delay Stats – There’s only one:
  - Both Implementations report (as loss ratio)
  - Type-P-One-way-Loss-Average <= revise to -Ratio
The ADK R-package returns some values and these require interpretation:

$t_i \text{obs}$ is calculated, an observed value based on an ADK metric. The absolute $t_i \text{obs}$ value must be less than or equal to the Critical Point.

The $P$-value or ($P$) in the following tables is a statistical test to bolster confidence in the result. It should be greater than or equal to $\alpha = 0.05$.

Critical Points for a confidence interval of 95% (or $\alpha = 0.05$)
For $k = 2$ samples, the Critical Point is 1.960
For $k = 4$ samples, the Critical Point is 1.915
For $k = 9$ samples, the Critical Point is 1.839
(Note, the ADK publication doesn't list a Critical Point for 8 samples, but it can be interpolated)

Green = ADK test passed, Red = ADK test failed
Percentiles of the ADK Criteria for various sample combinations (k= number of samples)
[Table 1 of Scholz and Stevens]

<table>
<thead>
<tr>
<th>m (k-1)</th>
<th>0.75 (\alpha=0.25)</th>
<th>0.90 (\alpha=0.1)</th>
<th>0.95 (\alpha=0.05)</th>
<th>0.975 (\alpha=0.025)</th>
<th>0.99 (\alpha=0.01)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.326</td>
<td>1.225</td>
<td>1.960</td>
<td>2.719</td>
<td>3.752</td>
</tr>
<tr>
<td>2</td>
<td>.449</td>
<td>1.309</td>
<td>1.945</td>
<td>2.576</td>
<td>3.414</td>
</tr>
<tr>
<td>3</td>
<td>.498</td>
<td>1.324</td>
<td>1.915</td>
<td>2.493</td>
<td>3.246</td>
</tr>
<tr>
<td>4</td>
<td>.525</td>
<td>1.329</td>
<td>1.894</td>
<td>2.438</td>
<td>3.139</td>
</tr>
</tbody>
</table>

Criteria met when |t.obs| < ADK Criteria(%-tile of interest)
Also: P-value should be > \(\alpha\) (rule of thumb)
Test Set-up Experiences

- Test bed set up may have to be described in more detail.
- We’ve worked with a single vendor.
- Selecting the proper Operation System took us one week (make sure support of L2TPv3 is a main purpose of that software).
- Connect the IPPM implementation to a switch and install a cable or internal U-turn on that switch. Maintain separate IEEE 802.1q logical VLAN connections when connecting the switch to the CPE which terminates the L2TPv3 tunnel.
- The CPE requires at least a route-able IP address as LB0 interface, if the L2TPv3 tunnel spans the Internet.
- The Ethernet Interface MUST be cross connected to the L2TPv3 tunnel in port mode.
- Terminate the L2TPv3 tunnel on the LB0 interface.
- Don’t forget to configure firewalls and other middle boxes properly.
NetProbe 5.8.5

- Runs on Solaris (and Linux, occasionally)
- Pre-dates *WAMP, functionally similar
- Software-based packet generator
- Provides performance measurements including Loss, Delay, PDV, Reordering, Duplication, burst loss, etc. in post-processing on stored packet records
Section 6.2 – Loss Threshold

- See Section 2.8.2 of [RFC2680].
- 1. configure a path with 1 sec one-way constant delay
- 2. measure (average) one-way delay with 2 or more implementations, using identical waiting time thresholds for loss set at 2 seconds
- 3. configure the path with 3 sec one-way delay (or change the delay while test is in progress, measurements in step 2)
- 4. repeat measurements
- 5. observe that the increase measured in step 4 caused all packets to be declared lost, and that all packets that arrive successfully in step 2 are assigned a valid one-way delay.