Advancing Metrics on the Standards Track:

RFC 2680 (1-way Loss) Test Plan and Results

draft-ietf-ippm-testplan-rfc2680-02

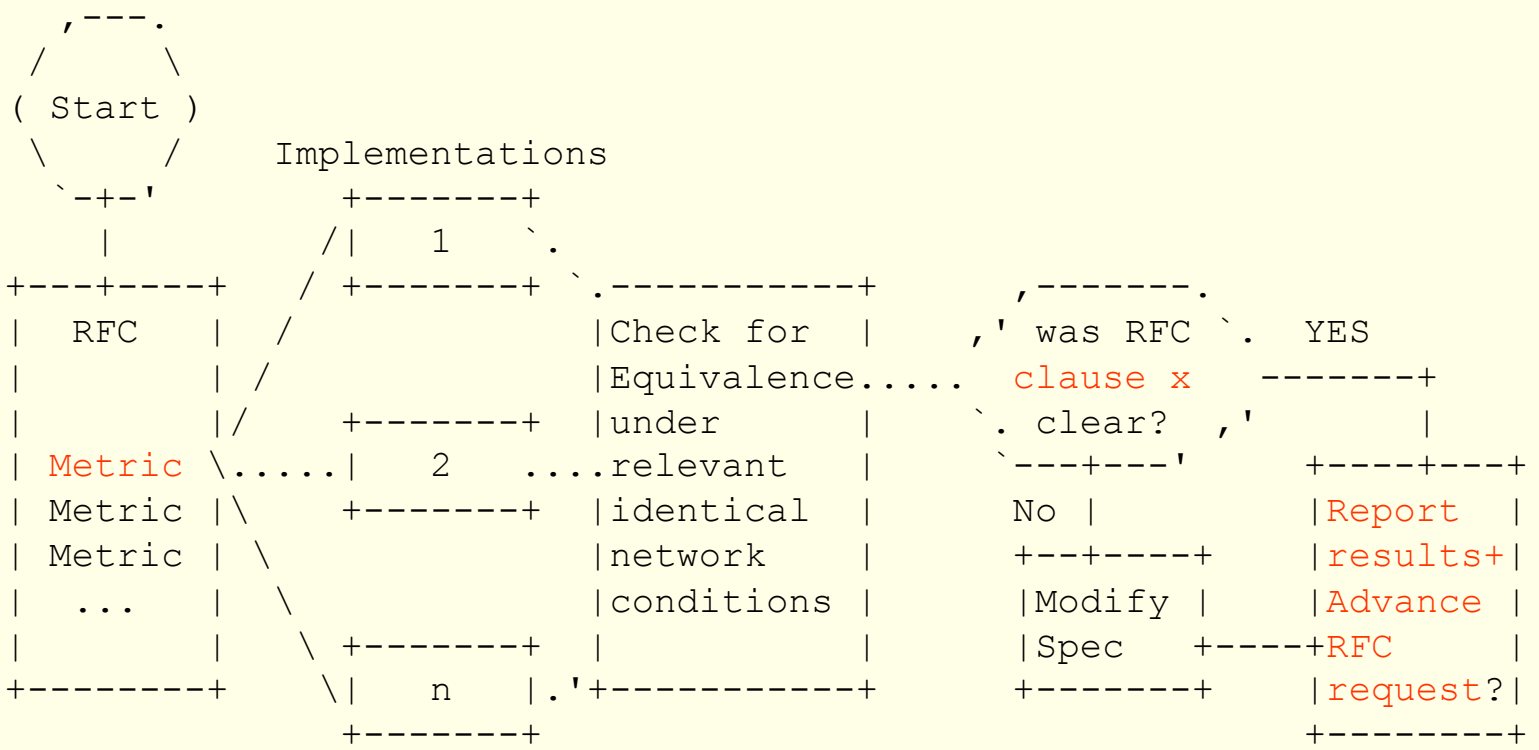
Len Ciavattone, Rüdiger Geib,
Al Morton, Matthias Wieser

March 2013

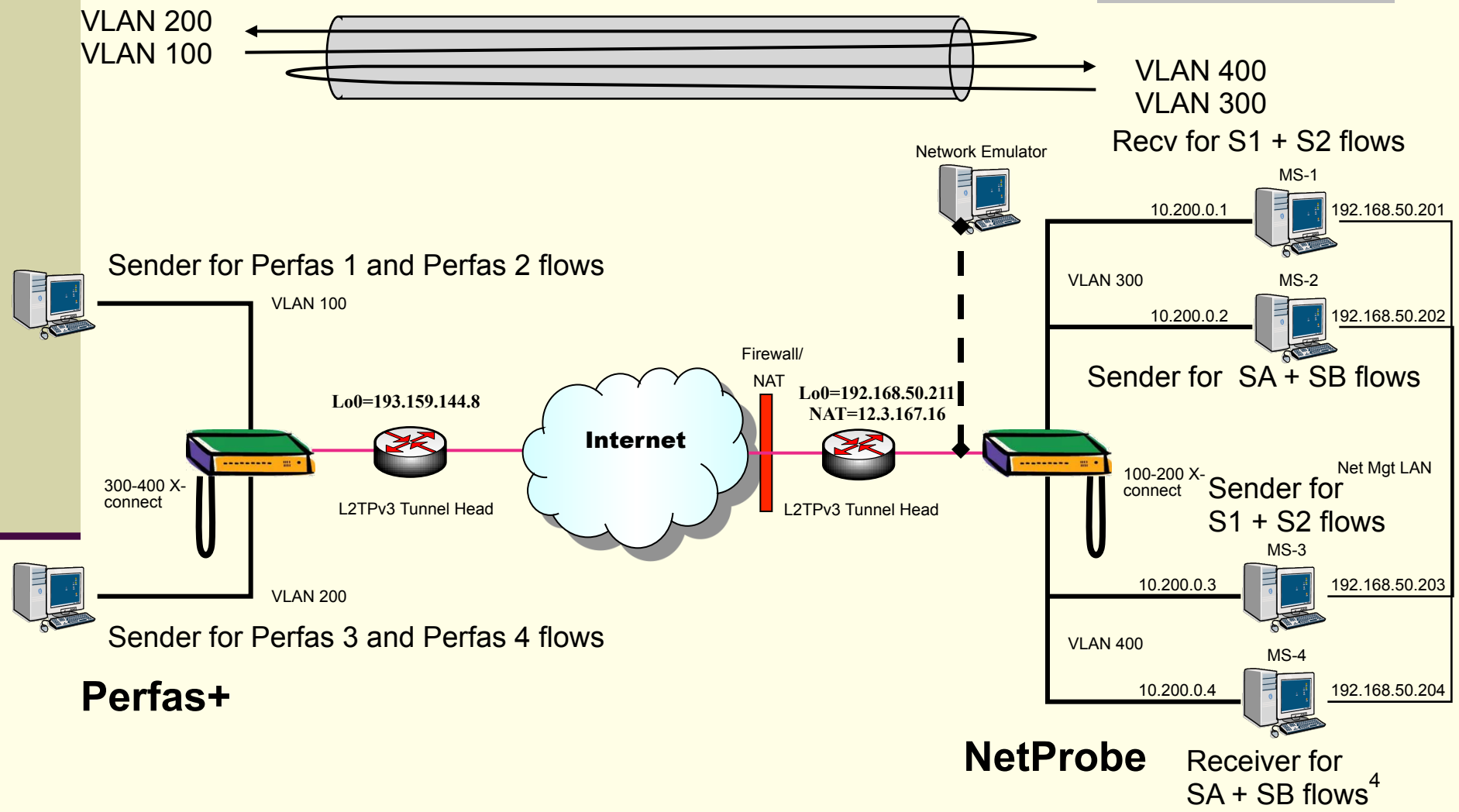
Outline

- 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



Test Configuration



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 distirbutions
  - 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

- R Development Core Team (2011), R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.
- Scholz F.W. and Stephens M.A. (1987), K-sample Anderson-Darling Tests, *Journal of the American Statistical Association,* **Vol 82, No. 399**, 918–924.

BACKUP

Backup Backup Backup

Overview of Testing (sample)

Date	Samp	Interval	Duration	Notes	ADK same	ADK cross
Mar 23	Poisson	1s	300s	Netem 10% Loss		
Mar 24	Periodic	1s	300s	Netem 100ms +/- 50ms delay		
Mar 24	Periodic	1s	300s	Netem 10% Loss		Pass
Mar 28	Periodic	1s	300s	Netem 100ms		
Mar 29	Periodic (rand st.)	1s	300s	Netem 100ms +/- 50ms delay, 64 Byte	NP s12AB Per p1234	Pass combined
Apr 6	Periodic (rand st.)	1s	300s	Netem 100ms +/- 50ms delay, 340 Byte		
Apr 7	Periodic (rand st.)	1s	1200s	Netem 10% Loss		Pass
Apr 12	Periodic (rand st.)	1s	300s	Netem 100ms, 500 Byte and 64 Byte comparison		

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.

340B 1s Periodic	**ti.obs**	**P-value***
not adj. for ties	0.52043	0.20604
adj. for ties	0.62679	0.18607
64B 1s Periodic		
not adj. for ties	0.76921	0.16200
adj. for ties	0.90935	0.14113
64B 1s Poisson**		
not adj. for ties	2.15099	0.04145
adj. for ties	1.93129	0.05125

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

ADK tests – Glossary & Background

The ADK R-package returns some values and these require interpretation:

ti.obs is calculated, an observed value based on an ADK metric. The absolute ti.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]

m (k-1)	0.75 α=0.25	0.90 α=0.1	0.95 α=0.05	0.975 α=0.025	0.99 α=0.01
1	.326	1.225	1.960	2.719	3.752
2	.449	1.309	1.945	2.576	3.414
3	.498	1.324	1.915	2.493	3.246
4	.525	1.329	1.894	2.438	3.139

Criteria met when |t.obs| < ADK Criteria(%-tile of interest)

Also: P-value should be > α (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.