Advancing Metrics on the Standards Track:

RFC2679 Test Plan

(and very soon, Results)

`draft-morton-ippm-testplan-rfc2679-00`

Len Ciavattone, Rüdiger Geib,
Al Morton, Matthias Wieser

March 2011

Outline

- Implement the Definition-centric metric advancement described in “metrictest” draft
- Test Plan Overview
  - Test Set-up and Specific Tests
- Key Discussion: Proposal on Thresholds of Equivalence (Anderson-Darling K-sample)
  - Key substitution for Interoperability
  - MUST be agreed in advance of results review
  - (and, since we will confirm any agreement on the list, so not done today)
- Qualitative description of testing in-progress

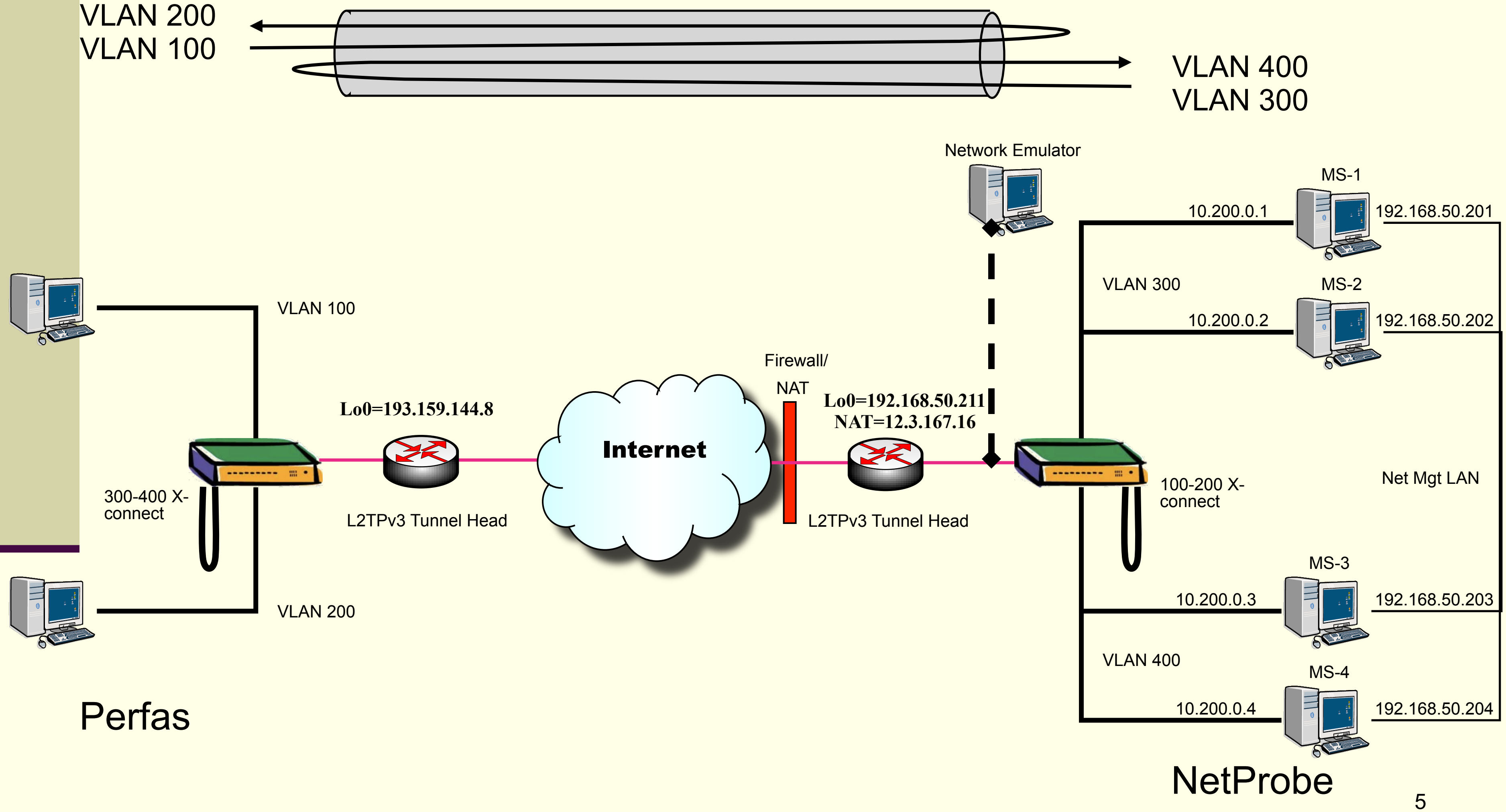
Definition-Centric Process

```
  ,---.
 /     \
( 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     |
+--------+    \|   n    |.'+------------+      +--------+    |request?|
               +--------+                                   +--------+
```

Key Points (the sub-points)

- Start with an RFC
  - Focus on a specific clause
- Run test(s) with Implementations
  - Test plan is customized to a specific clause
- Evaluate Measurements & Compare
  - Expected measurement results are Clear
  - Obvious place to take action if text is found to be ambiguous
- Final state is Report Dev. for Protocol Action Req.

Test Configuration



Tests in the Plan

- 6. Tests to evaluate RFC 2679 Specifications
  - 6.1. One-way Delay, ADK Sample Comparison – Same & Cross Implementations
  - 6.2. One-way Delay, Loss threshold,
  - 6.3. One-way Delay, First-bit to Last bit,
  - 6.4. One-way Delay, Difference Sample Metric
  - 6.5. Implementation of Statistics for One-way Delay

Section 6.1 One-way Delay, ADK Sample Comparisons (Same/Cross)

1. Configure tests on an L2TPv3 tunnel over a live network path.
2. Measure a sample of one-way delay singletons with 2 or more implementations, using identical options.
3. Measure a sample of one-way delay singletons with *four* instances of the *same* implementations,
   - connectivity differences SHOULD be the same as for the *cross* implementation tests.
4. Apply ADK comparison: same (see App C of metrictest)
5. Take coarsest confidence/resolution, or Section 5 Limits
6. Apply constant correction factors (Section 5)
7. Compare Cross-Implementation ADK for equivalence (samples come from same distribution)

Decide Equivalence Limits First!

- Through the now-fixed publication of our proposal on limits, the Test Team has effectively provided a way to move forward, with two possible outcomes:
  - IPPM reaches CONSENSUS on the limits and allowances in section 5 (before any results are published).
  - IPPM REVISES the limits and allowances WITHOUT input from the test team (who are just now looking at the data collected successfully)
    - except to clarify the details of the testing and set-up, and reaches consensus on the new limits.
- In either case, we compare the results with the agreed limits at some future time
  - Test Team had hoped that would be *now*

Proposal for the Equivalence Threshold and Correction Factors

- Need to AGREE on these Criteria before evaluating the results (e.g.,VLAN test set-up has non-identical path components)
- Purpose: Evaluate Specification Clarity (using results implementations)
- For ADK comparison: cross-implementations
  - 0.95 confidence factor at 1ms resolution, or
  - The smallest confidence factor & res. of *same* Imp.
- A constant time accuracy error < +/-0.5ms MAY be removed from one Implementation before ADK or comparison of means
- A constant propagation delay error < +2ms MAY be removed from one Implementation …
  - (due to use of different sub-nets between the switch and measurement devices at each location)

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.

Brief Overview of Testing In-Progress

- Difficulties achieving communications on the test set-up.
  - VLANs
  - (new) Network Emulator(s)
- Close inspection = info on implementations
- Preliminary testing and Statistical findings
  - We have seen many cases of successful comparison, but also cases where the comparisons failed and we are working to understand the factors that influence the outcomes.

Summary

- Test Plan for Key clauses of RFC 2679
  - Would be the basis of Advance RFC Request
  - Should this be a WG document?
- Two Implementations: NetProbe and Perfas
- Experiments begun
- Proposal for Equivalence Threshold and Correction Factors ← Need Consensus!

BACKUP

Backup Backup Backup

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 3.5 of [RFC2679], 3rd bullet point and also Section 3.8.2 of [RFC2679].
- 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.

Section 6.3: First-bit to Last-bit

See Section 3.7.2 of [RFC2679], and Section 10.2 of [RFC2330].

- 1. configure a path with 1000 ms one-way constant delay, and ideally including a low-speed link (10-baseT, FD)
- 2. measure (average) one-way delay with 2 or more implementations, using identical options and equal size small packets (e.g., 32 octet IP payload)
- 3. maintain the same path with 1000 ms one-way delay
- 4. measure (average) one-way delay with 2 or more implementations, using identical options and equal size large packets (e.g., 1400 octet IP payload)
- 5. observe that the increase measured in steps 2 and 4 is equivalent to the increase in ms expected due to the larger serialization time for each implementation. Most of the measurement errors in each system should cancel, if they are stationary.

Other Examples

- 6.4 One-way Delay, RFC 2679
  - This test is intended to evaluate measurements in sections 3 and 4 of [RFC2679].

    Average delays before/after 2 second increase

- 4. Error Calibration, RFC 2679
  - This is a simple check to determine if an implementation reports the error calibration as required in Section 4.8 of [RFC2679].