Workgroup: Network Working Group
Internet-Draft:
draft-morton-ippm-capacity-metric-protocol-02
Published: 25 October 2021
Intended Status: Standards Track
Expires: 28 April 2022
Authors: L. Ciavattone AT&T Labs
A. Morton AT&T Labs

Test Protocol for One-way IP Capacity Measurement

Abstract

This memo addresses the problem of protocol support for measuring Network Capacity metrics in RFC xxxx: draft-ietf-ippm-capacity-metric-method, where the method deploys a feedback channel from the receiver to control the sender's transmission rate in near-real-time. It supplies a simple protocol to perform the measurements.

See Section 10: The authors seek feedback to determine what additional features will be necessary for an IETF Standards Track Protocol, beyond what is present in the running code available now.

Status of This Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at https://datatracker.ietf.org/drafts/current/.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on 28 April 2022.

Copyright Notice

Copyright (c) 2021 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (https://trustee.ietf.org/license-info) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with

respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

1. Introduction
  1.1. Requirements Language
2. Scope, Goals, and Applicability
3. Protocol Overview
4. General Parameters and Definitions
5. Setup Request and Response Exchange
  5.1. Setup Response Processing at the Client
6. Test Activation Request and Response
  6.1. Test Activation Request at the client
  6.2. Test Activation Request - server response
  6.3. Test Activation Response action at the client
7. Test Stream Transmission and Measurement Feedback Messages
  7.1. Test Packet PDU and Roles
  7.2. Status PDU
8. Stopping the Test
9. Method of Measurement
  9.1. Running Code
10. Security Considerations
11. IANA Considerations
12. Acknowledgments
13. References
  13.1. Normative References
  13.2. Informative References
Authors' Addresses

1. Introduction

The IETF's efforts to define Network and Bulk Transport Capacity have been chartered and finally progressed after over twenty years.

Over that time, the performance community has seen development of Informative definitions in [RFC3148] for Framework for Bulk Transport Capacity (BTC), RFC 5136 for Network Capacity and Maximum IP-layer Capacity, and the Experimental metric definitions and methods in [RFC8337], Model-Based Metrics for BTC.

This memo looks at the problem of measuring Network Capacity metrics defined in [I-D.ietf-ippm-capacity-metric-method] where the method deploys a feedback channel from the receiver to control the sender's transmission rate in near-real-time.

Although there are several test protocol already available for support and manage active measurements, this protocol is a major departure from their operation:

1. UDP transport is used for all setup, test activation, and control messages, and for results feedback (not TCP), simplifying operations.

2. TWAMP [RFC5357] and STAMP [RFC8762] use the philosophy that one host is a Session-Reflector, sending test packets every time they receive a test packet. This protocol supports a one-way test with periodic status messages returned to the sender. These messages are also a basis for on-path Round-trip delay measurements, which are a key input to the load adjustment search algorithm.

3. OWAMP [RFC4656] supports one-way testing with results Fetch at the end of the test session. This protocol supports a one-way test and requires periodic status messages returned to the sender to support the load adjustment search algorithm.

4. The security features of OWAMP [RFC4656] and TWAMP [RFC5357] have been described as "unusual", to the point that IESG approved their use while also asking that these methods not be used again. Further, the common OWAMP [RFC4656] and TWAMP [RFC5357] approach to security is over 15 years old at this time.

Note: the -02 update of this draft will be the last that describes version 8 of the protocol in the running code. Future updates of the draft will correspond to protocol version 9 and higher versions.

1.1. Requirements Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "NOT RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in BCP 14[RFC2119] [RFC8174] when, and only when, they appear in all capitals, as shown here.

2. Scope, Goals, and Applicability

The scope of this memo is to define a protocol to measure the Maximum IP-Layer Capacity metric and according to the standardized method.

The continued goal is to harmonize the specified metric and method across the industry, and this protocol supports the specifications of IETF and other Standards Development Organizations.

All active testing protocols currently defined by the IPPM WG are UDP-based, but this protocol specifies both control and test protocols using UDP transport. Also, the control protocol continues operating during testing to convey results and dynamic configurations.

The primary application of the protocol described here is the same as in Section 2 of [RFC7497] where:

*The access portion of the network is the focus of this problem statement. The user typically subscribes to a service with bidirectional access partly described by rates in bits per second.

3. Protocol Overview

This section gives an informative overview of the communication protocol between two test end-points (without expressing requirements: later sections provide details and requirements).

One end-point takes the role of server, awaiting connection requests on a well-known port from the other end-point, the client.

The client requires configuration of a test direction parameter (upstream or downstream test) as well as the hostname or IP address of the server in order to begin the setup and configuration exchanges with the server.

The protocol uses UDP transport and has four phases:

1. Setup Request and Response Exchange: The client requests to begin a test by communicating its protocol version, intended security mode, and jumbo datagram support. The server either confirms matching configuration or rejects the connection. The server also communicates the ephemeral port for further communication when accepting the client's request.

2. Test Activation Request and Response: the client composes a request conveying parameters such as the testing direction, the duration of the test interval and test sub-intervals, and various thresholds. The server then chooses to accept, ignore or modify any of the test parameters, and communicates the set that will be used unless the client rejects the modifications. Note that the client assumes that the Test Activation exchange has opened any co-located firewalls and network address/port translators for the test connection (in response to the Request packet on the ephemeral port) and the traffic that follows. If the Test Activation is rejected or fails, the client assumes that the firewall will close the address/port combination after the firewall's configured idle traffic time-out.

3. Test Stream Transmission and Measurement Feedback Messages: Testing proceeds with one end-point sending load PDUs and the other end-point receiving the load PDUs and sending frequent status messages to communicate status and transmission conditions there. The feedback messsages are input to a load-control algorithm at the server, which controls future sending rates at either end-point as needed. The choice to locate the load-control algorithm at the server, regardlesss of transmiision direction, means that the algorithm can be updated more easily at a host within the network, and at a fewer number of hosts than the number of clients.

4. Stopping the Test: The server initiates the phase to stop the test by setting the STOP1 indication in load PDUs or sstatus feedback messages. The client acknowledges by setting the STOP2 in further load PDUs or messages, and a graceful connection termination at each end-point follows. (Since the load PDUs and feedback messages are used, this phase is kind of a sub-phase of 3.) If the Test traffic stops or the communication path fails, the client assumes that the firewall will close the address/port combination after the firewall's configured idle traffic time-out.

4. General Parameters and Definitions

This section lists the REQUIRED input factors to specify a Sender or Receiver metric.

*Src, the address of a host (such as the globally routable IP address).

*Dst, the address of a host (such as the globally routable IP address).

*MaxHops, the limit on the number of Hops a specific packet may visit as it traverses from the host at Src to the host at Dst (implemented in the TTL or Hop Limit).

*T0, the time at the start of measurement interval, when packets are first transmitted from the Source.

*I, the nominal duration of a measurement interval at the destination (default 10 sec)

*dt, the nominal duration of m equal sub-intervals in I at the destination (default 1 sec)

*dtn, the beginning boundary of a specific sub-interval, n, one of m sub-intervals in I

*FT, the feedback time interval between status feedback messages communicating measurement results, sent from the receiver to control the sender. The results are evaluated to determine how to adjust the current offered load rate at the sender (default 50ms)

*Tmax, a maximum waiting time for test packets to arrive at the destination, set sufficiently long to disambiguate packets with long delays from packets that are discarded (lost), such that the distribution of one-way delay is not truncated.

*F, the number of different flows synthesized by the method (default 1 flow)

*flow, the stream of packets with the same n-tuple of designated header fields that (when held constant) result in identical treatment in a multi-path decision (such as the decision taken in load balancing). Note: The IPv6 flow label MAY be included in the flow definition when routers have complied with [RFC6438] guidelines.

*Type-P, the complete description of the test packets for which this assessment applies (including the flow-defining fields). Note that the UDP transport layer is one requirement for test packets specified below. Type-P is a parallel concept to "population of interest" defined in clause 6.1.1 of[Y.1540].

*PM, a list of fundamental metrics, such as loss, delay, and reordering, and corresponding target performance threshold. At least one fundamental metric and target performance threshold MUST be supplied (such as One-way IP Packet Loss [RFC7680] equal to zero).

A non-Parameter which is required for several metrics is defined below:

*T, the host time of the *first* test packet's *arrival* as measured at the destination Measurement Point, or MP(Dst). There may be other packets sent between source and destination hosts that are excluded, so this is the time of arrival of the first packet used for measurement of the metric.

Note that time stamp format and resolution, sequence numbers, etc. will be established by this memo.

5. Setup Request and Response Exchange

All messages defined in this section SHALL use UDP transport. The hosts SHALL calculate and include the UDP checksum, or check the UDP checksum as neccessary.

The client SHALL begin the Control protocol connection by sending a Setup Request message to the server's control port.

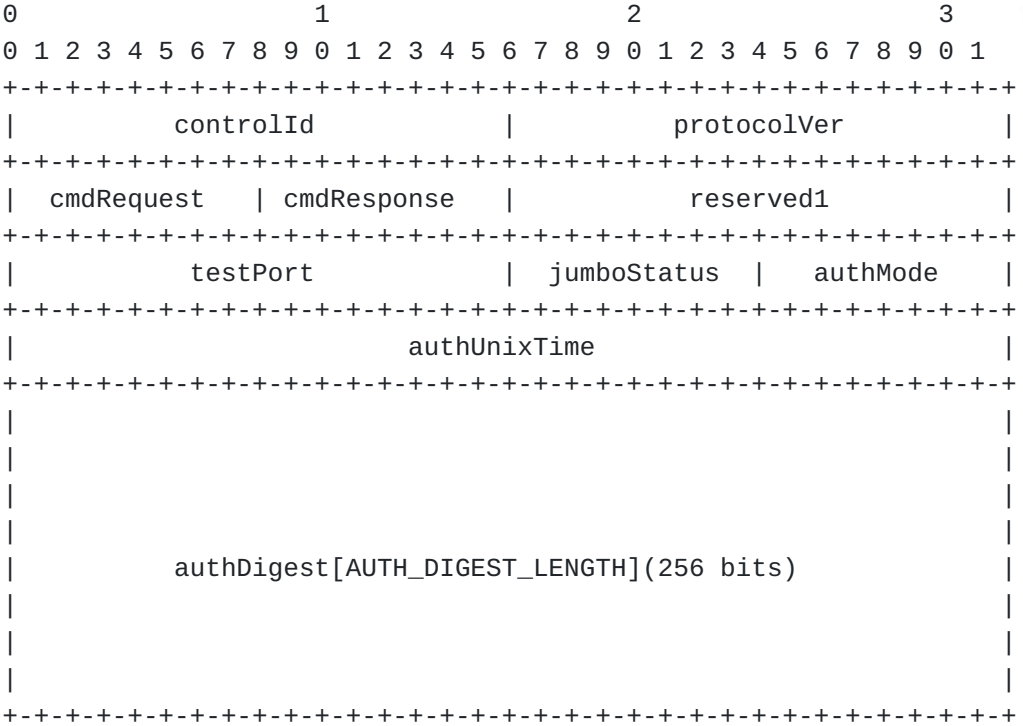
The client SHALL simultaneously start a test initiation timer so that if the control protocol fails to complete all exchanges in the allocated time, the client software SHALL exit (close the UDP socket and indicate an error message to the user).

(Note: in version 8, the watchdog time-out is configured, in udpst.h, as #define WARNING_NOTRAFFIC 1 // Receive traffic stopped warning threshold (sec) #define TIMEOUT_NOTRAFFIC (WARNING_NOTRAFFIC + 4) or 5 seconds)

The Setup Request message PDU SHALL be organized as follows:

```
uint16_t controlId;   // Control ID = 0xACE1
uint16_t protocolVer; // Protocol version = 0x08
uint8_t cmdRequest;   // Command request = 1 (request)
uint8_t cmdResponse;  // Command response = 0
uint16_t reserved1;   // Reserved (alignment)
uint16_t testPort;    // Test port on server  (=0 for Request)
uint8_t jumboStatus;  // Jumbo datagram support status (BOOL)
uint8_t authMode;     // Authentication mode
uint32_t authUnixTime;// Authentication time stamp
unsigned char authDigest[AUTH_DIGEST_LENGTH] // SHA256_DIGEST_LE
```

The UDP PDU format layout SHALL be as follows (big-endian AB):



When the server receives the Setup Request it SHALL validate the request by checking the protocol version, the jumbo datagram support indicator, and the authentication data if utilized. If the client has selected options for:

*Jumbo datagram support status (BOOL),

*Authentication mode, and

*Authentication time stamp

that do not match the server configuration, the server MUST reject the Setup Request.

(Note: in version 8, the watchdog time is configured, in udpst.h, as #define WARNING_NOTRAFFIC 1 // Receive traffic stopped warning threshold (sec) #define TIMEOUT_NOTRAFFIC (WARNING_NOTRAFFIC + 4) or 5 seconds)

If the Setup Request must be rejected (due to any of the reasons in the Command response codes listed below), a Setup Response SHALL be sent back to the client with a corresponding command response value indicating the reason for the rejection.

```
uint16_t controlId;   // Control ID = 0xACE1
uint16_t protocolVer; // Protocol version = 0x08
uint8_t cmdRequest;   // Command request = 2 (reply)
uint8_t cmdResponse;  // Command response = <see table below>
uint16_t reserved1;   // Reserved (alignment)
uint16_t testPort;    // Test port on server  (available port in
uint8_t jumboStatus;  // Jumbo datagram support status (BOOL)
uint8_t authMode;     // Authentication mode
uint32_t authUnixTime;// Authentication time stamp
unsigned char authDigest[AUTH_DIGEST_LENGTH] // 32 octets, MBZ
```

```
Command Response Codes
Control Header Setup Request Code CHSR_CRSP_NONE     0 = None
Control Header Setup Request Code CHSR_CRSP_ACKOK    1 = Acknowledgement
Control Header Setup Request Code CHSR_CRSP_BADVER   2 = Bad Protocol Ve
Control Header Setup Request Code CHSR_CRSP_BADJS    3 = Invalid Jumbo d
Control Header Setup Request Code CHSR_CRSP_AUTHNC   4 = Unexpected Auth
Control Header Setup Request Code CHSR_CRSP_AUTHREQ  5 = Authentication
Control Header Setup Request Code CHSR_CRSP_AUTHINV  6 = Invalid authent
Control Header Setup Request Code CHSR_CRSP_AUTHFAIL 7 = Authentication
Control Header Setup Request Code CHSR_CRSP_AUTHTIME 8 = Authentication
```

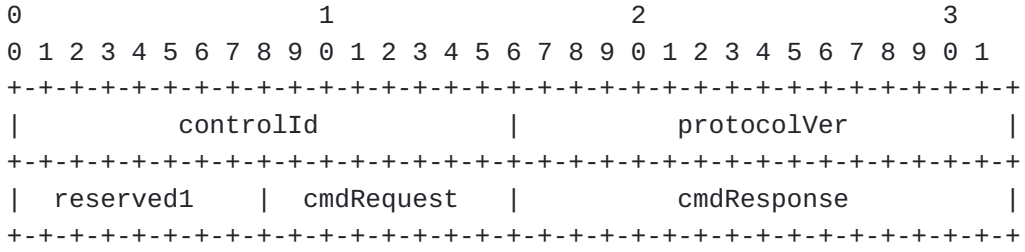
@@@@ To Do: How do we communicate multiple errors when the server sends the Setup Response? Is an error hierarchy possible, where Bad Protocol Version means that none of the other aspects (higher error numbers) were checked? New text to address this issue appears below:

There is a hierarchy of Command Response codes, beginning with: "2 = Bad Protocol Version", which SHALL be checked first because it is a fixed field length and the most reliable check. The server SHOULD communicate the first error condition detected in the order listed below:

3 = Invalid Jumbo datagram option
5 = Authentication missing in Setup Request
4 = Unexpected Authentication in Setup Request
6 = Invalid authentication method (SHA-256 not used)
7 = Authentication failure (both shared secret and time)
8 = Authentication time is invalid in Setup Request (replay attack)

The only circumstance when a server would not communicate the appropriate Command Response Code for an error condition above is when an attack has been detected, in which case the server will allow setup attempts with errors to terminate silently.

@@@@ Or, if multiple error codes are wanted, would a flag system work better? For an expanded field multiple error codes, we could use decimal values that set 1 bit in the cmdResponse (0,1,2,4,8,...) for each code, and expand the cmdResponse field to 16 bits by using the nearby reserved field as shown below (and moving the fields within the 32-bit word):



@@@@ - end text for discussion -

If the server finds that the Setup Request matches its configuration and is otherwise acceptable, the server SHALL initiate a new connection for the client, using a new UDP socket allocated from the UDP ephemeral port range. Then, the server SHALL start a watchdog timer (to terminate the connection in case the client goes silent), and sends the Setup Response back to the client (see below for composition).

If the Setup Request is accepted by the server, a Setup Response SHALL be sent back to the client with a corresponding command response value indicating 1 = Acknowledgement.

```
uint16_t controlId;   // Control ID = 0xACE1
uint16_t protocolVer; // Protocol version = 0x08
uint8_t cmdRequest;   // Command request = 2 (reply)
uint8_t cmdResponse;  // Command response = 1 (Acknowledgement)
uint16_t reserved1;   // Reserved (alignment)
uint16_t testPort;    // Test port on server  (available port in
uint8_t jumboStatus;  // Jumbo datagram support status (BOOL)
uint8_t authMode;     // Authentication mode
uint32_t authUnixTime;// Authentication time stamp
unsigned char authDigest[AUTH_DIGEST_LENGTH] // 32 octets, MBZ
...
```

The new connection is associated with a new UDP socket allocated from the UDP ephemeral port range at the server. The server SHALL set a timer for the new connection as a watchdog (in case the client goes quiet) and send the Setup response back to the client.

(Note: in version 8, the watchdog time-out is configured at 5 seconds)

The Setup Response SHALL include the port number at the server for the new socket, and this UDP port-pair SHALL be used for all subsequent communication. The server SHALL also include the values of:

- *Jumbo datagram support status (BOOL),
- *Authentication mode, and
- *Authentication time stamp

for the client's use on the new connection in its Setup Response, and the remaining 32 octets MUST Be Zero (MBZ).

Finally, the new UDP connection associated with the new socket and port number is opened, and the server awaits communication there.

If a Test Activation request is not subsequently received from the client on this new port number before the watchdog timer expires, the server SHALL close the socket and deallocate the port.

5.1. Setup Response Processing at the Client

When the client receives the Setup response from the server it first checks the cmdResponse value. If this value indicates an error the client SHALL display/report a relevant message to the user or management process and exit. If the client receives a Command Response code (CRSP) that is not equal to one of the codes defined above, then the client MUST terminate the connection and terminate operation of the current Setup Request. If the Command Response code

(CRSP) value indicates success the client SHALL compose a Test Activation Request with all the test parameters it desires such as the test direction, the test duration, etc.

6. Test Activation Request and Response

This section is divided according to the sending and processing of the client, server, and again at the client.

All messages defined in this section SHALL use UDP transport. The hosts SHALL calculate and include the UDP checksum, or check the UDP checksum as neccessary.

6.1. Test Activation Request at the client

Upon a successful setup, the client SHALL then send the Test Activation Request to the UDP port number the server communicated in the Setup Response.

@@@@ To Do: Add Options for UDP payload content (beyond the Test PDU), such as all zeroes, all ones, alternating one and zero, and pseudo-random.

The client SHALL compose Test Activation Request as follows:

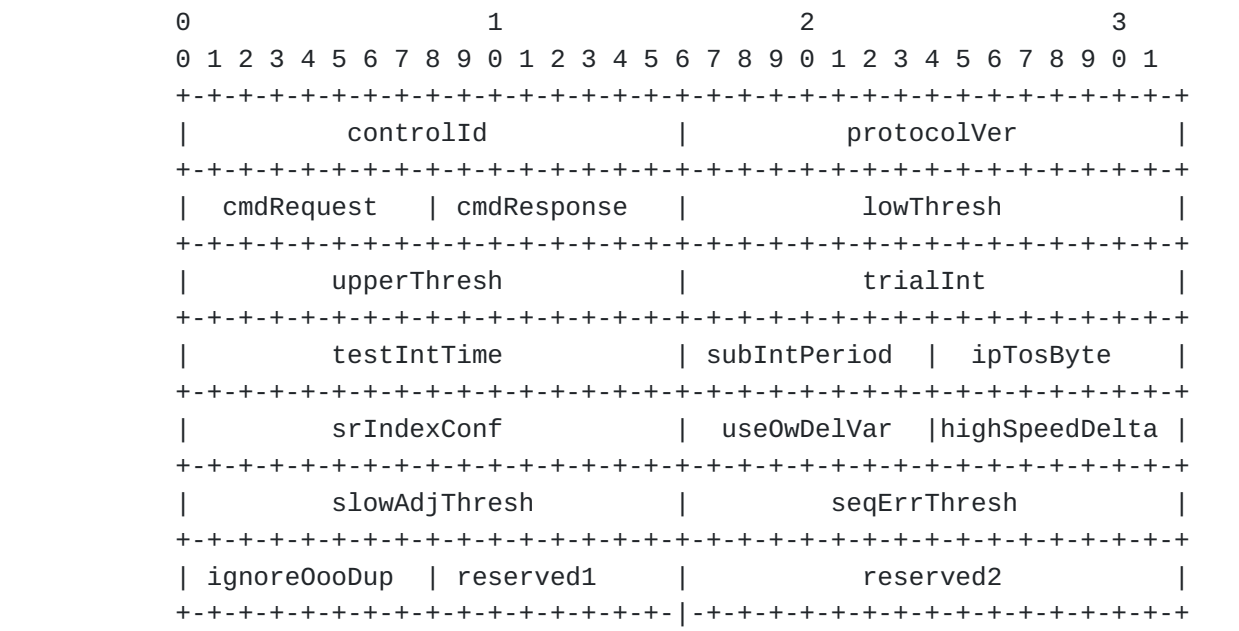
```
        uint16_t controlId;   // Control ID
        uint16_t protocolVer; // Protocol version
        uint8_t cmdRequest; // Command request, 1 = upstream, 2 = downst
        uint8_t cmdResponse;         // Command response (set to 0)
        uint16_t lowThresh;          // Low delay variation threshold
        uint16_t upperThresh;        // Upper delay variation threshold
        uint16_t trialInt;           // Status feedback/trial interval (
        uint16_t testIntTime;        // Test interval time (sec)
        uint8_t subIntPeriod;        // Sub-interval period (sec)
        uint8_t ipTosByte;           // IP ToS byte for testing
        uint16_t srIndexConf;        // Configured sending rate index (s
        uint8_t useOwDelVar;         // Use one-way delay instead of RTT
        uint8_t highSpeedDelta;      // High-speed row adjustment delta
        uint16_t slowAdjThresh;      // Slow rate adjustment threshold
        uint16_t seqErrThresh;       // Sequence error threshold
        uint8_t ignoreOooDup;        // Ignore Out-of-Order/Duplicate da
        uint8_t reserved1;           // (Alignment)
        uint16_t reserved2;          // (Alignment)

Control Header Test Activation Command Request Values:
CHTA_CREQ_NONE      0 = No Request
CHTA_CREQ_TESTACTUS 1 = Request test in Upstream direction (client to se
CHTA_CREQ_TESTACTDS 2 = Request test in Downstream direction (server to

Control Header Test Activation Command Response Values:
CHTA_CRSP_NONE     0 = Used by client when making a Request
CHTA_CRSP_ACKOK    1 = Used by Server in affirmative Response
CHTA_CRSP_BADPARAM 2 = Used by Server to indicate an error; bad paramete
```

Note: uint16_t srIndexConf is the table index of the configured *fixed* sending rate index to use. The client can request the specified rate, or the server can use this field to coerce a maximum rate in its response. If the server sets to 0 in its response, client SHALL not use fixed rate.

The UDP PDU format of the Test Activation Request is as follows (big-endian AB):



Note: This is only 28 octets of the 56 octet PDU sent, the rest are MBZ for a Test Activation Request.

The client SHALL use the configuration for

- *Jumbo datagram support status (BOOL),
- *Authentication mode, and
- *Authentication time stamp

requested and confirmed by the server.

6.2. Test Activation Request - server response

After the server receives the Test Activation request on the new connection, it MUST choose to accept, ignore or modify any of the test parameters.

When the server sends the Test Activation response back, it SHALL set the cmd Response field to:

uint8_t cmdResponse;// Command response (set to 1, ACK, or 2 error)

The server SHALL include all the test parameters again to make the client aware of any changes.

If the client has requested an upstream test, the server SHALL include the transmission parameters from the first row of the sending rate table.

The remaining 28 octets of the Test Activation Request (normally read from the first row of the sending rate table) are called the Sending Rate Structure, and SHALL be organized as follows:

```
     uint32_t txInterval1; // Transmit interval (us)
     uint32_t udpPayload1; // UDP payload (bytes)
     uint32_t burstSize1;  // UDP burst size per interval
     uint32_t txInterval2; // Transmit interval (us)
     uint32_t udpPayload2; // UDP payload (bytes)
     uint32_t burstSize2;  // UDP burst size per interval
     uint32_t udpAddon2;   // UDP add-on (bytes)
```

with



Note that the server additionally has the option of completely rejecting the request and sending back an appropriate command response value:

```
uint8_t cmdResponse; // Command response (set to 2, error)
```

If activation continues, the new connection is prepared for an upstream OR downstream test.

In the case of a downstream test, the server prepares to send with either a single timer to send status PDUs at the specified interval OR dual timers to send load PDUs based on the first row of the sending rate table.

The server SHALL then send a Test Activation response back to the client, update the watchdog timer with a new time-out value, and set a test duration timer to eventually stop the test.

The new connection is now ready for testing.

6.3. Test Activation Response action at the client

When the client receives the Test Activation response, it first checks the command response value.

If the client receives a Test Activation command response value that indicates an error, the client SHALL display/report a relevant message to the user or management process and exit.

If the client receives a Test Activation command response value that is not equal to one of the codes defined above, then the client MUST terminate the connection and terminate operation of the current Setup Request.

If the client receives a Test Activation command response value that indicates success (CHTA_CRSP_ACKOK) the client SHALL update its configuration to use any test parameters modified by the server.

Next, the client SHALL prepare its connection for either an upstream test with dual timers set to send load PDUs (based on the starting transmission parameters sent by the server), OR a downstream test with a single timer to send status PDUs at the specified interval.

Then, the client SHALL stop the test initiation timer, set a new time-out value for the watchdog timer, and start the timer (in case the server goes quiet).

The connection is now ready for testing.

7. Test Stream Transmission and Measurement Feedback Messages

This section describes the testing phase of the protocol. The roles of sender and receiver vary depending whether the direction of testing is from server to client, or the reverse.

All messages defined in this section SHALL use UDP transport. The hosts SHALL calculate and include the UDP checksum, or check the received UDP checksum before further processing, as neccessary.

7.1. Test Packet PDU and Roles

Testing proceeds with one end point sending load PDUs, based on transmission parameters from the sending rate table, and the other end point receiving the load PDUs and sending status messages to communicate the traffic conditions at the receiver.

The watchdog timer at the receiver SHALL be reset each time a test PDU is received. See non-graceful test stop in Section 8 for handling the watchdog/NOTRAFFIC time-out expiration at each end-point.

When the server is sending Load PDUs in the role of sender, it SHALL use the transmission parameters directly from the sending rate table via the index that is currently selected (which was based on the feedback in its received status messages).

However, when the client is sending load PDUs in the role of sender, it SHALL use the discreet transmission parameters that were communicated by the server in its periodic status messages (and not referencing a sending rate table). This approach allows the server to control the individual sending rates as well as the algorithm used to decide when and how to adjust the rate.

The server uses a load adjustment algorithm which evaluates measurements, either it's own or the contents of received feedback messages. This algorithm is unique to udpst; it provides the ability to search for the Maximum IP Capacity that is absent from other testing tools. Although the algorithm depends on the protocol, it is not part of the protocol per se.

The current algorithm has three paths to its decision on the next sending rate:

1. When there are no impairments present (no sequence errors, low delay variation), resulting in sending rate increase.

2. When there are low impairments present (no sequence errors but higher levels of delay variation), so the same sending rate is retained.

3. When the impairment levels are above the thresholds set for this purpose and "congestion" is inferred, resulting in sending rate decrease.

The algorithm also has two modes for increasing/decreasing the sending rate:

*A high-speed mode to achieve high sending rates quickly, but also back-off quickly when "congestion" is inferred from the measurements. Any two consecutive feedback intervals that have a sequence number anomaly and/or contain an upper delay variation threshold exception in both of the two consecutive intervals, count as the two consecutive feedback measurements required to declare "congestion" within a test.

*A single-step mode where all rate adjustments use the minimum increase or decrease of one step in the sending rate table. The single step mode continues after the first inference of "congestion" from measured impairments.

On the other hand, the test configuration MAY use a fixed sending rate requested by the client, using the field below:

```
uint16_t srIndexConf; // Configured sending rate index
```

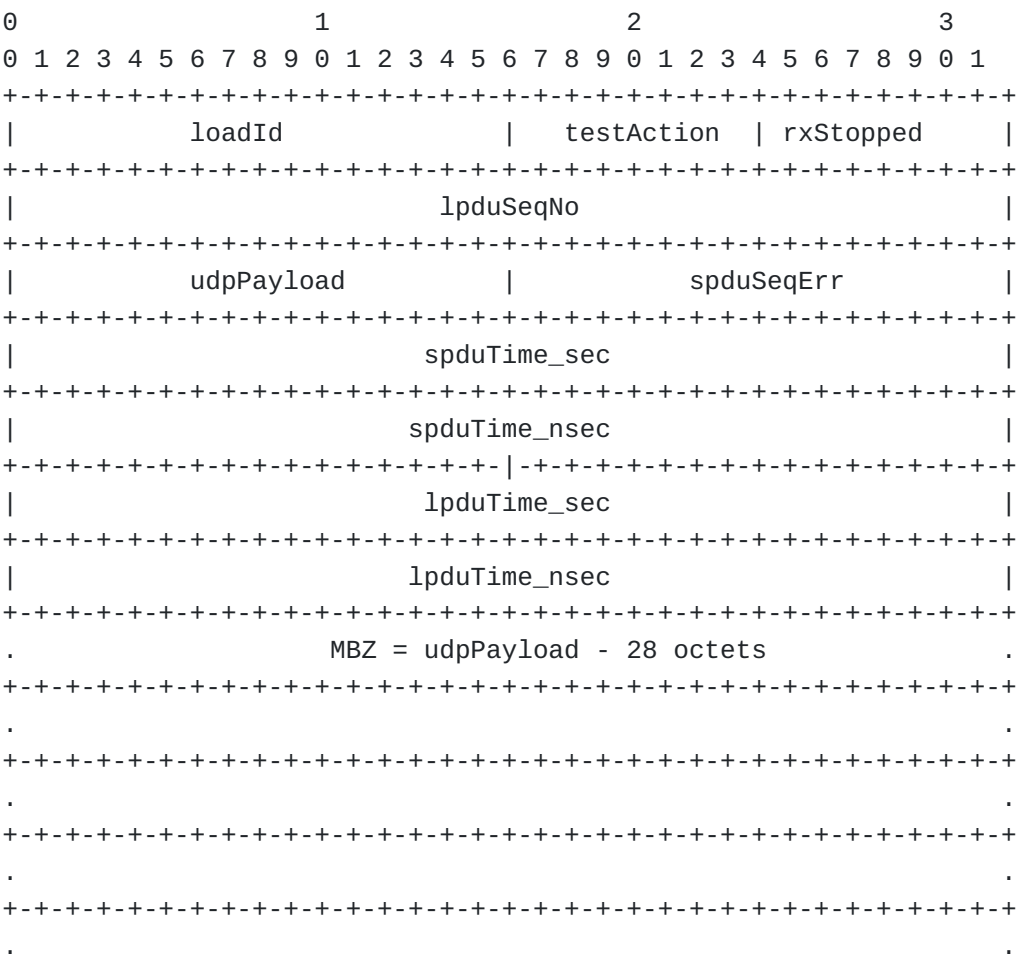
The client MAY communicate the desired fixed rate in it's activation request.

The Load PDU SHALL have the following format and field definitions:

```
     uint16_t loadId; // Load ID (=0xBEEF for the LOad PDU)
     uint8_t testAction;  // Test action (= 0x00 normally, until test
     uint8_t rxStopped;   // Receive traffic stopped indicator (BOOL)
     uint32_t lpduSeqNo;  // Load PDU sequence number (starts at 1)
     uint16_t udpPayload; // UDP payload LENGTH(bytes)
     uint16_t spduSeqErr; // Status PDU sequence error count
     //
     uint32_t spduTime_sec;  // Send time in last received status PDU
     uint32_t spduTime_nsec; // Send time in last received status PDU
     uint32_t lpduTime_sec;  // Send time of this load PDU
     uint32_t lpduTime_nsec; // Send time of this load PDU
```

```
Test Action Codes
TEST_ACT_TEST  0  // normal
TEST_ACT_STOP1 1  // normal stop at end of test: server sends in STATUS
TEST_ACT_STOP2 2  // ACK of STOP1: sent by client in STATUS or Test PDU
```

The Test Load UDP PDU format is as follows (big-endian AB):



7.2. Status PDU

The receiver SHALL send a Status PDU to the sender during a test at the configured feedback interval.

The watchdog timer at the test PDU sender SHALL be reset each time a Status PDU is received. See non-graceful test stop in Section 8 for handling the watchdog/NOTRAFFIC time-out expiration at each end-point.

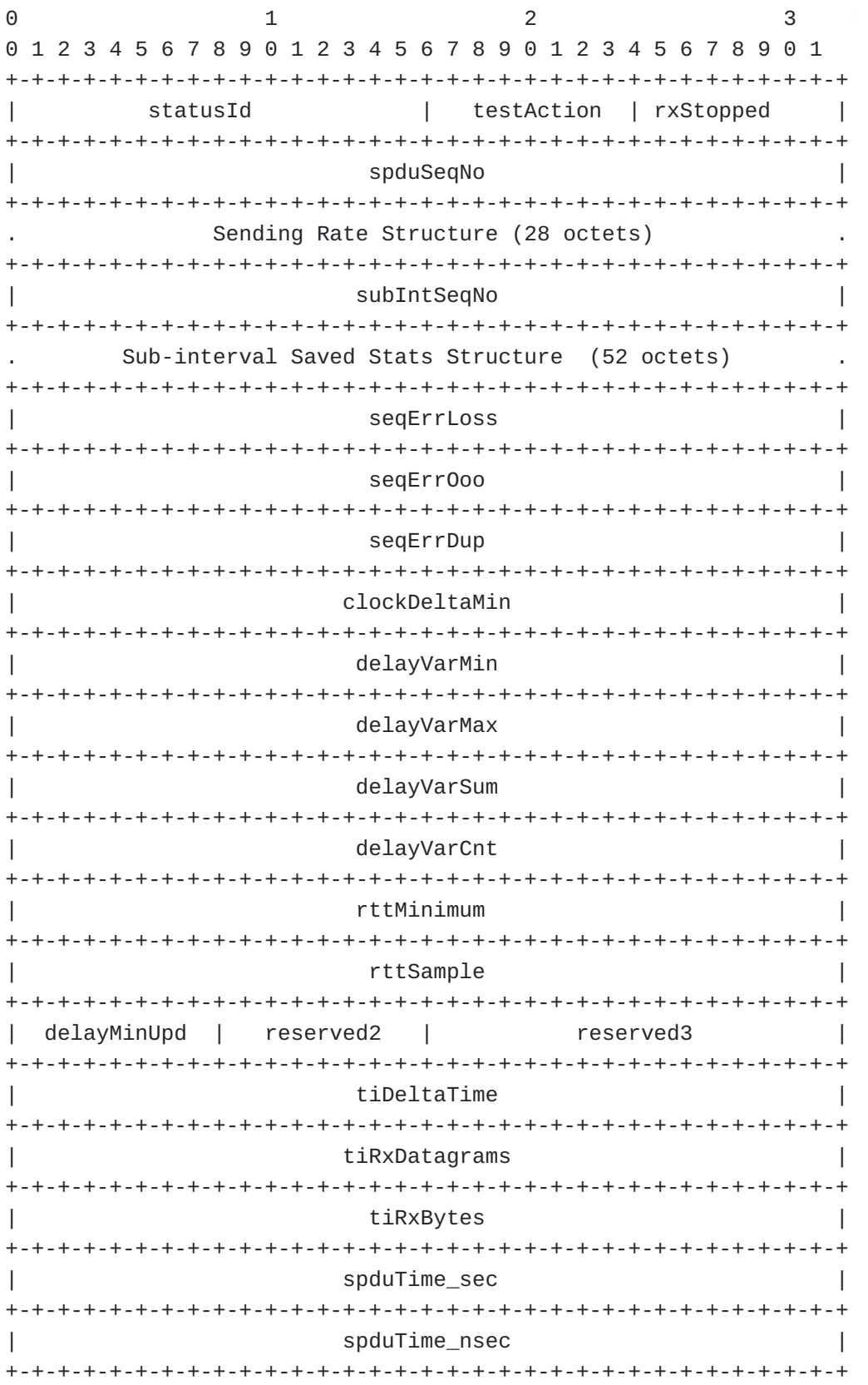
@@@@ To Do: What protections from bit errors (checksum) or on-path attacks (something stronger) are warrented for teh Status PDUs? These PDUs are a key part of the server-client control loop. Added a requirement to calculate and include/check the UDP checksum.

The Status Header PDU SHALL have the following format and field definitions:

```
// Status feedback header for UDP payload of status PDUs
//

        uint16_t statusId;  // Status ID = 0xFEED
        uint8_t testAction; // Test action
        uint8_t rxStopped;  // Receive traffic stopped indicator (BOOL)
        uint32_t spduSeqNo; // Status PDU sequence number (starts at 1)
        //
        struct sendingRate srStruct; // Sending Rate Structure (28 octet
        //
        uint32_t subIntSeqNo;      // Sub-interval sequence number
        struct subIntStats sisSav; // Sub-interval Saved Stats Structure
        //
        uint32_t seqErrLoss; // Loss sum
        uint32_t seqErrOoo;  // Out-of-Order sum
        uint32_t seqErrDup;  // Duplicate sum
        //
        uint32_t clockDeltaMin; // Clock delta minimum (either RTT or 1-
        uint32_t delayVarMin;   // Delay variation minimum
        uint32_t delayVarMax;   // Delay variation maximum
        uint32_t delayVarSum;   // Delay variation sum
        uint32_t delayVarCnt;   // Delay variation count
        uint32_t rttMinimum;    // Minimum round-trip time sampled
        uint32_t rttSample;     // Last round-trip time sample
        uint8_t delayMinUpd;    // Delay minimum(s) updated observed, co
        uint8_t reserved2;      // (alignment)
        uint16_t reserved3;     // (alignment)
        //
        uint32_t tiDeltaTime;   // Trial interval delta time
        uint32_t tiRxDatagrams; // Trial interval receive datagrams
        uint32_t tiRxBytes;     // Trial interval receive bytes
        //
        uint32_t spduTime_sec;  // Send time of this status PDU
        uint32_t spduTime_nsec; // Send time of this status PDU
```

The Status feedback UDP payload PDUs format is as follows (big-endian AB):



Note that the Sending Rate Structure (28 octets) is defined in the Test Activation section.

Also note that the Sub-interval Saved Stats Structure (52 octets) SHALL be included (and populated as required when the server is in the receiver role) as defined below.

The Sub-interval saved statistics structure for received traffic measurements SHALL be organized and formatted as follows:

```
      uint32_t rxDatagrams; // Received datagrams
      uint32_t rxBytes;     // Received bytes
      uint32_t deltaTime;   // Time delta
      uint32_t seqErrLoss;  // Loss sum
      uint32_t seqErrOoo;   // Out-of-Order sum
      uint32_t seqErrDup;   // Duplicate sum
      uint32_t delayVarMin; // Delay variation minimum
      uint32_t delayVarMax; // Delay variation maximum
      uint32_t delayVarSum; // Delay variation sum
      uint32_t delayVarCnt; // Delay variation count
      uint32_t rttMinimum;  // Minimum round-trip time
      uint32_t rttMaximum;  // Maximum round-trip time
      uint32_t accumTime;   // Accumulated time
```

```
   0                   1                   2                   3
   0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                           rxDatagrams                         |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                            rxBytes                            |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                           deltaTime                           |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                           seqErrLoss                          |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                           seqErrOoo                           |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                           seqErrDup                           |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                           delayVarMin                         |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                           delayVarMax                         |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                           delayVarSum                         |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                           delayVarCnt                         |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                           rttMinimum                          |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                           rttMaximum                          |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                           accumTime                           |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Note that the 52 octet saved statistics structure above has slight differences from the 40 octets that follow in the status feedback PDU, particularly the time-related fields.

Upon receiving the Status Feedback PDU or expiration of the feedback interval, the server SHALL perform calculations required by the Load adjustment algorithm and adjust its sending rate, or signal that the client do so in its role as as sender.

@@@@ To Do: Additional measurements, like interface byte counters from a client at a residential gateway, would change the Status Feedback PDU (and the protocol version number as a result). Interface byte counters seem useful for specific circumstances, such as when the client application has acces to an interface that sees all traffic to/from a service subscriber's location.

8. Stopping the Test

When the test duration timer on the server expires, it SHALL set the connection test action to STOP and also starts marking all outgoing load or status PDUs with a test action of STOP1.

uint8_t testAction; // Test action (server sets STOP1)

This is simply a non-reversible state for all future messages sent from the server.

When the client receives a load or status PDU with the STOP1 indication, it SHALL finalize testing, display the test results, and also mark its connection with a test action of STOP (so that any PDUs received subsequent to the STOP1 are ignored).

With the test action of the client's connection set to STOP, the very next expiry of a send timer for either a load or status PDU SHALL cause the client to schedule an immediate end time to exit.

The client SHALL then send all subsequent load or status PDUs with a test action of STOP2

uint8_t testAction; // Test action (client sets STOP2)

as confirmation to the server, and a graceful termination of the test can begin.

When the server receives the STOP2 confirmation in the load or status PDU, the server SHALL schedule an immediate end time for the connection which closes the socket and deallocates it.

In a non-graceful test stop, the watchdog/NOTRAFFIC time-outs at each end-point will expire (sometimes at one end-point first), notifications in logs, STDOUT, and/or formateed output SHALL be made, and the test action of each end-point's connection SHALL be set to STOP.

9. Method of Measurement

The architecture of the method REQUIRES two cooperating hosts operating in the roles of Src (test packet sender) and Dst (receiver), with a measured path and return path between them.

The duration of a test duration, parameter I, MUST be constrained in a production network, since this is an active test method and it will likely cause congestion on the Src to Dst host path during a test.

9.1. Running Code

This section is for the benefit of the Document Shepherd's form, and will be deleted prior to final review.

Much of the development of the method and comparisons with existing methods conducted at IETF Hackathons and elsewhere have been based on the example udpst Linux measurement tool (which is a working reference for further development) [udpst]. The current project:

- *is a utility that can function as a client or server daemon

- *requires a successful client-initiated setup handshake between cooperating hosts and allows firewalls to control inbound unsolicited UDP which either go to a control port [expected and w/authentication] or to ephemeral ports that are only created as needed. Firewalls protecting each host can both continue to do their job normally. This aspect is similar to many other test utilities available.

- *is written in C, and built with gcc (release 9.3) and its standard run-time libraries

- *allows configuration of most of the parameters described in Sections 4 and 7.

- *supports IPv4 and IPv6 address families.

- *supports IP-layer packet marking.

10. Security Considerations

Active metrics and measurements have a long history of security considerations. The security considerations that apply to any active measurement of live paths are relevant here. See [RFC4656] and [RFC5357].

When considering privacy of those involved in measurement or those whose traffic is measured, the sensitive information available to

potential observers is greatly reduced when using active techniques which are within this scope of work. Passive observations of user traffic for measurement purposes raise many privacy issues. We refer the reader to the privacy considerations described in the Large Scale Measurement of Broadband Performance (LMAP) Framework [RFC7594], which covers active and passive techniques.

There are some new considerations for Capacity measurement as described in this memo.

1. Cooperating source and destination hosts and agreements to test the path between the hosts are REQUIRED. Hosts perform in either the Src or Dst roles.

2. It is REQUIRED to have a user client-initiated setup handshake between cooperating hosts that allows firewalls to control inbound unsolicited UDP traffic which either goes to a control port [expected and w/authentication] or to ephemeral ports that are only created as needed. Firewalls protecting each host can both continue to do their job normally.

3. Client-server authentication and integrity protection for feedback messages conveying measurements is RECOMMENDED. To accomodate different host limitations and testing circumstances, different modes of operation are recommended:

   A. Unathenticated mode (for all phases)
   AND
   B. OPTIONAL Authenticated set-up only
   SHA-256 HMAC time-window verification (5 min time stamp verification)
   (could add silent failure option)

   -=-=-=-=-=-=-=-=-=-

   C. Encrypted setup and test-activation
   (currently using OpenSSL Library, so KISS, but may be too slow for test packets)

   -=-=-=-=--=- Old/lowpower host performance impacts -=-=-=-=-=-=-

   D. Encrypted feedback messages (maybe split into Integrity and encrypt?

   E. Integrity protection for test packets SHA-256 HMAC

   F. Encrypted test packets (maybe also valuable to defeat compression on

4. Hosts MUST limit the number of simultaneous tests to avoid resource exhaustion and inaccurate results.

5. Senders MUST be rate-limited. This can be accomplished using a pre-built table defining all the offered load rates that will be supported (Section 8.1). The recommended load-control search algorithm results in "ramp up" from the lowest rate in the table.

6. Service subscribers with limited data volumes who conduct extensive capacity testing might experience the effects of Service Provider controls on their service. Testing with the Service Provider's measurement hosts SHOULD be limited in frequency and/or overall volume of test traffic (for example, the range of I duration values SHOULD be limited).

The exact specification of these features was hopefully accomplished during this protocol development.

11. IANA Considerations

This memo requests IANA to assign a UDP port.

12. Acknowledgments

Thanks to Ruediger Geib, Lincoln Lavoie, Can Desem, and Greg Mirsky for reviewing this draft and providing helpful suggestions and areas for further development.

13. References

13.1. Normative References

[I-D.ietf-ippm-capacity-metric-method] Morton, A., Geib, R., and L. Ciavattone, "Metrics and Methods for One-way IP Capacity", Work in Progress, Internet-Draft, draft-ietf-ippm-capacity-metric-method-12, 9 June 2021, <https://www.ietf.org/archive/id/draft-ietf-ippm-capacity-metric-method-12.txt>.

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <https://www.rfc-editor.org/info/rfc2119>.

[RFC2330] Paxson, V., Almes, G., Mahdavi, J., and M. Mathis, "Framework for IP Performance Metrics", RFC 2330, DOI 10.17487/RFC2330, May 1998, <https://www.rfc-editor.org/info/rfc2330>.

[RFC2681] Almes, G., Kalidindi, S., and M. Zekauskas, "A Round-trip Delay Metric for IPPM", RFC 2681, DOI 10.17487/RFC2681,

September 1999, <https://www.rfc-editor.org/info/rfc2681>.

[RFC6438] Carpenter, B. and S. Amante, "Using the IPv6 Flow Label for Equal Cost Multipath Routing and Link Aggregation in Tunnels", RFC 6438, DOI 10.17487/RFC6438, November 2011, <https://www.rfc-editor.org/info/rfc6438>.

[RFC7497] Morton, A., "Rate Measurement Test Protocol Problem Statement and Requirements", RFC 7497, DOI 10.17487/RFC7497, April 2015, <https://www.rfc-editor.org/info/rfc7497>.

[RFC7680] Almes, G., Kalidindi, S., Zekauskas, M., and A. Morton, Ed., "A One-Way Loss Metric for IP Performance Metrics (IPPM)", STD 82, RFC 7680, DOI 10.17487/RFC7680, January 2016, <https://www.rfc-editor.org/info/rfc7680>.

[RFC8174] Leiba, B., "Ambiguity of Uppercase vs Lowercase in RFC 2119 Key Words", BCP 14, RFC 8174, DOI 10.17487/RFC8174, May 2017, <https://www.rfc-editor.org/info/rfc8174>.

[RFC8468] Morton, A., Fabini, J., Elkins, N., Ackermann, M., and V. Hegde, "IPv4, IPv6, and IPv4-IPv6 Coexistence: Updates for the IP Performance Metrics (IPPM) Framework", RFC 8468, DOI 10.17487/RFC8468, November 2018, <https://www.rfc-editor.org/info/rfc8468>.

13.2. Informative References

[copycat] Edleine, K., Kuhlewind, K., Trammell, B., and B. Donnet, "copycat: Testing Differential Treatment of New Transport Protocols in the Wild (ANRW '17)", 15 July 2017, <https://irtf.org/anrw/2017/anrw17-final5.pdf>.

[LS-SG12-A] 12, I. S., "LS - Harmonization of IP Capacity and Latency Parameters: Revision of Draft Rec. Y.1540 on IP packet transfer performance parameters and New Annex A with Lab Evaluation Plan", May 2019, <https://datatracker.ietf.org/liaison/1632/>.

[LS-SG12-B] 12, I. S., "LS on harmonization of IP Capacity and Latency Parameters: Consent of Draft Rec. Y.1540 on IP packet transfer performance parameters and New Annex A with Lab & Field Evaluation Plans", March 2019, <https://datatracker.ietf.org/liaison/1645/>.

[RFC2544] Bradner, S. and J. McQuaid, "Benchmarking Methodology for Network Interconnect Devices", RFC 2544, DOI 10.17487/

RFC2544, March 1999, <https://www.rfc-editor.org/info/rfc2544>.

[RFC3148] Mathis, M. and M. Allman, "A Framework for Defining Empirical Bulk Transfer Capacity Metrics", RFC 3148, DOI 10.17487/RFC3148, July 2001, <https://www.rfc-editor.org/info/rfc3148>.

[RFC4656] Shalunov, S., Teitelbaum, B., Karp, A., Boote, J., and M. Zekauskas, "A One-way Active Measurement Protocol (OWAMP)", RFC 4656, DOI 10.17487/RFC4656, September 2006, <https://www.rfc-editor.org/info/rfc4656>.

[RFC5136] Chimento, P. and J. Ishac, "Defining Network Capacity", RFC 5136, DOI 10.17487/RFC5136, February 2008, <https://www.rfc-editor.org/info/rfc5136>.

[RFC5357] Hedayat, K., Krzanowski, R., Morton, A., Yum, K., and J. Babiarz, "A Two-Way Active Measurement Protocol (TWAMP)", RFC 5357, DOI 10.17487/RFC5357, October 2008, <https://www.rfc-editor.org/info/rfc5357>.

[RFC6815] Bradner, S., Dubray, K., McQuaid, J., and A. Morton, "Applicability Statement for RFC 2544: Use on Production Networks Considered Harmful", RFC 6815, DOI 10.17487/RFC6815, November 2012, <https://www.rfc-editor.org/info/rfc6815>.

[RFC7312] Fabini, J. and A. Morton, "Advanced Stream and Sampling Framework for IP Performance Metrics (IPPM)", RFC 7312, DOI 10.17487/RFC7312, August 2014, <https://www.rfc-editor.org/info/rfc7312>.

[RFC7594] Eardley, P., Morton, A., Bagnulo, M., Burbridge, T., Aitken, P., and A. Akhter, "A Framework for Large-Scale Measurement of Broadband Performance (LMAP)", RFC 7594, DOI 10.17487/RFC7594, September 2015, <https://www.rfc-editor.org/info/rfc7594>.

[RFC7799] Morton, A., "Active and Passive Metrics and Methods (with Hybrid Types In-Between)", RFC 7799, DOI 10.17487/RFC7799, May 2016, <https://www.rfc-editor.org/info/rfc7799>.

[RFC8337] Mathis, M. and A. Morton, "Model-Based Metrics for Bulk Transport Capacity", RFC 8337, DOI 10.17487/RFC8337, March 2018, <https://www.rfc-editor.org/info/rfc8337>.

[RFC8762] Mirsky, G., Jun, G., Nydell, H., and R. Foote, "Simple Two-Way Active Measurement Protocol", RFC 8762, DOI

10.17487/RFC8762, March 2020, <https://www.rfc-editor.org/info/rfc8762>.

[RFC8972] Mirsky, G., Min, X., Nydell, H., Foote, R., Masputra, A., and E. Ruffini, "Simple Two-Way Active Measurement Protocol Optional Extensions", RFC 8972, DOI 10.17487/RFC8972, January 2021, <https://www.rfc-editor.org/info/rfc8972>.

[TR-471] Morton, A., "Broadband Forum TR-471: IP Layer Capacity Metrics and Measurement", 10 July 2020, <https://www.broadband-forum.org/technical/download/TR-471.pdf>.

[udpst] udpst Project Collaborators, "UDP Speed Test Open Broadband project", December 2020, <https://github.com/BroadbandForum/obudpst>.

[Y.1540] Y.1540, I. R., "Internet protocol data communication service - IP packet transfer and availability performance parameters", December 2019, <https://www.itu.int/rec/T-REC-Y.1540-201912-I/en>.

[Y.Sup60] Morton, A., Rapporteur., "Recommendation Y.Sup60, (09/20) Interpreting ITU-T Y.1540 maximum IP-layer capacity measurements", 11 September 2020, <https://www.itu.int/rec/T-REC-Y.Sup60/en>.

Authors' Addresses

Len Ciavattone
AT&T Labs
200 Laurel Avenue South
Middletown,, NJ 07748
United States of America

Email: lencia@att.com

Al Morton
AT&T Labs
200 Laurel Avenue South
Middletown,, NJ 07748
United States of America

Phone: +1 732 420 1571
Email: acm@research.att.com