Network Working Group
Internet-Draft
Intended status: Informational
Expires: September 22, 2016

G. Fioccola
Telecom Italia
A. Clemm
Cisco Systems
M. Cociglio
Telecom Italia
M. Chandramouli
Cisco Systems
A. Capello
Telecom Italia

March 21, 2016

Alternate Marking Extension to Cisco SLA Protocol RFC6812 draft-fioccola-ippm-rfc6812-alt-mark-ext-01

Abstract

Cisco's Service-Level Assurance Protocol (Cisco's SLA Protocol) RFC 6812 [RFC6812] is a Performance Measurement protocol that has been widely deployed. The protocol is used to measure service-level parameters such as network latency, delay variation, and packet/frame loss. This document describes an extension to the Cisco SLA Protocol Measurement-Type UDP-Measurement, in order to implement alternate marking methodology detailed in [I-D.tempia-ippm-p3m]. The extension is used to measure service level parameters by marking test traffic.

Requirements Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC 2119 [RFC2119].

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 http://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 September 22, 2016.

Table of Contents

1. Introduction

Cisco SLA Protocol involves a system sending synthetic test traffic, which is reflected by a responder back to the sender. In the course, both sender and responder add a set of time stamps to the packet.

The packet is first time stamped when it is sent. A second time stamp is added by responder when it is received, and third time stamp when packet is sent back. A fourth time stamp is added when the sender receives the reflected packet. Based on time stamps and other information, the sender computes performance metrics such as loss, delay, and jitter.

One technique for passive performance measurements is described in [I-D.tempia-ippm-p3m]. This technique involves marking production flows as they traverse the network, then analyzing flow data associated with those marked flows. Passive measurements are very accurate in that they measure actual production traffic. However, there are scenarios in which passive measurements are not an option. For example, there may be no suitable flows currently occurring between pairs of nodes to be measured, or traffic may be tunneled and not be accessible to marking. In such cases, active measurements using synthetic test traffic need to be considered.

This document specifies an extension to Cisco SLA Protocol which allows to use Cisco SLA Protocol to generate synthetic traffic, but allows subjecting test traffic to the same techniqe described in [I-D.tempia-ippm-p3m]. Instead of time stamping test traffic, test traffic is marked and measurements occur by analyzing resulting flow data.

2. Description of the method

In order to perform packet loss, delay and jitter measurements on a traffic flow, different approaches exist. The method proposed consists in counting and timestamping the packets sent from one end, the packets received on the other end, and compare the two values. Therefore the devices performing the measurement have to refer exactly to the same set of packets. So the flow is virtually spit in consecutive blocks by coloring the packets so that the packets belonging to the same block will have the same color, whilst consecutive blocks will have different colors. Each change of color represents a sort of auto-synchronization signal that guarantees the consistency of measurements taken by different devices along the path.

This approach, called Alternate Marking method, is efficient both for passive performance monitoring and for active performance monitoring. In this document we describe the implementation for Active Measurement.

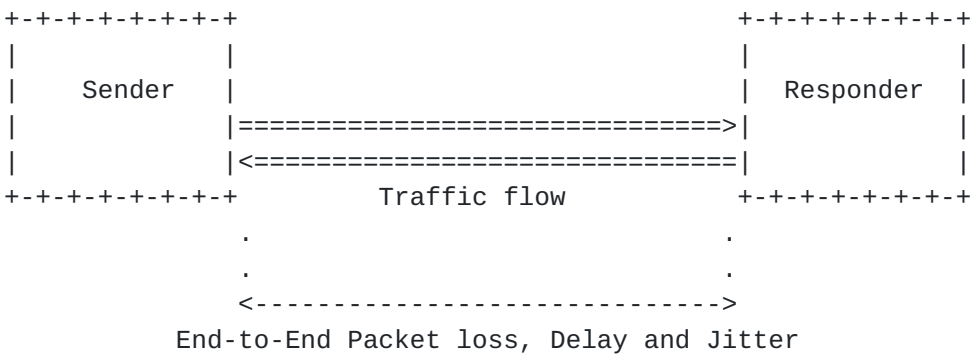


Figure 1: Available measurements

Previous Figure represents two end points (Sender and Responder) that exchange two equal data flows in both direction. The data flows start and end together. Packets are colored and the color changes every marking interval. The method can be used to measure packet loss, delay and jitter.

2.1. Packet loss measurement

The basic idea is to virtually split traffic flows into consecutive blocks: each block represents a measurable entity unambiguously recognizable by all network devices along the path. By counting the number of packets in each block and comparing the values measured by Sender and Responder, it is possible to measure packet loss occurred in any single block between the two end points.

A simple way to create the blocks is to "color" the traffic (two colors are sufficient) so that packets belonging to different consecutive blocks will have different colors. Whenever the color changes, the previous block terminates and the new one begins. The number of packets in each block depends on the criterion used to create the blocks: if the color is switched after a fixed number of packets, then each block will contain the same number of packets (except for any losses); but if the color is switched according to a fixed timer, then the number of packets may be different in each block depending on the packet rate.

2.2. Delay measurement

The same principle used to measure packet loss can be applied also to one-way delay measurement. There are two alternatives, shown below:

- Delay for each packet: For active measurement two alternate marking data flows are generated in both direction, so the alternation of colors can be used as a time reference to calculate the delay. Whenever the color changes (that means that a new

block has started) an end point can store the timestamps of all packets of the new block. The timestamps can be compared with the timestamps of the same packets on the other end point to compute packet delay. This method for measuring the delay is sensitive to out of order reception of packets. In order to overcome this problem between packets there should be a security time gap to avoid out of order issues. If the packet rate exchanged between the two end points is adequate each end points can store all the timestamp of the block and the packet delay can be computed for all the packets of the block, included minimum, maximum, average and median delay values.

- Average delay: A different approach, based on the concept of average delay, can be take in account for active measurement. The average delay is calculated by considering the average arrival time of the packets within a single block. The network device locally stores a timestamp for each packet received within a single block: summing all the timestamps and dividing by the total number of packets received, the average arrival time for that block of packets can be calculated. By subtracting the average arrival times of the two end points it is possible to calculate the average delay. This method is robust to out of order packets and also to packet loss (only a small error is introduced). Moreover, it greatly reduces the number of timestamps (only one per block for each end point) that have to be collected and transmitted for the calculation. On the other hand, it only gives one measure for the duration of the block, and it doesn't give the minimum, maximum and median delay values. RFC 6703 [RFC6703] recommends to report both median and average delay in order to obtain additional information about the distribution. But the same procedure of the average delay is not applicable to median delay because the median is not a linear operator. So the average delay could be considered as a light measurement because the calculation is achieved by exchanging only the average timestamp for each colored block (without exchanging all the timestamps). For this reason the average delay is not the main technique, but a secondary option in case we have to save computational resources.

By summing the one-way delay measurements of the two directions of a path, it is also possible to measure the two-way delay (round-trip delay).

In brief, there are three choices to compute delay for active measurement:

- The two end points could store all packets timestamps in both directions. At the end of the period all timestamps are exchanged. In this way, delay is calculated for each packet.

- The two end points calculate only the average timestamp that is exchanged at the end of the period. In this way only the average delay is calculated.

- The two end points sent packets with a specified and shared traffic profile and each end point could make its own calculation (data are not exchanged so it is not so accurate, but it depends on hardware and software capabilities).

Note: How data and timestamps are exchanged is outside the scope of this document.

2.3. Delay variation measurement

Similarly to one-way delay measurement, the method can also be used to measure the inter-arrival jitter. The alternation of colors can be used as a time reference to measure delay variations. The inter-arrival jitter can be easily derived from one-way delay measurement, by evaluating the delay variation of consecutive samples.

The concept of average delay can also be applied to delay variation, by evaluating the variation of average interval between consecutive packets of the flow.

3. Hybrid measurement

In order to have both end to end measurements and intermediate measurements (hybrid measurements) Sender and Responder exchanges traffic flows and apply alternate marking over these flows. In the intermediate points artificial traffic is managed in the same way as real traffic and measured as specified before.

4. Protocol

The Alternate Marking extension to Cisco Service Level Assurance Protocol consists of three distinct phases, Control Phase, Measurement Phase and Calculation Phase.

The Control Phase is the first phase of message exchanges and forms the base protocol. This phase establishes the identity of the Sender and provides information for the Measurement Phase. A single message pair of Control Request and Control Response marks this phase. The Sender initiates a Control Request message that is acknowledged by the Responder with a Control Response message. The Control Request may be sent multiple times if a Control Response has not been received; the number of times the message is retried is configurable on the Sender element.

The Measurement Phase forms the second phase and is comprised of an exchange of two equal alternate marking data flows between Sender and Responder. Sender and Responder generate test traffic and apply marking, not traffic is reflected and no timestamping is added to packets.

The Calculation Phase is introduced "ad hoc" for Alternate Marking implementation because it does not exist in Cisco Service Level Assurance Protocol described in RFC 6812 [RFC6812]. After test execution there are some alternatives to compute packet loss, delay and delay variation:

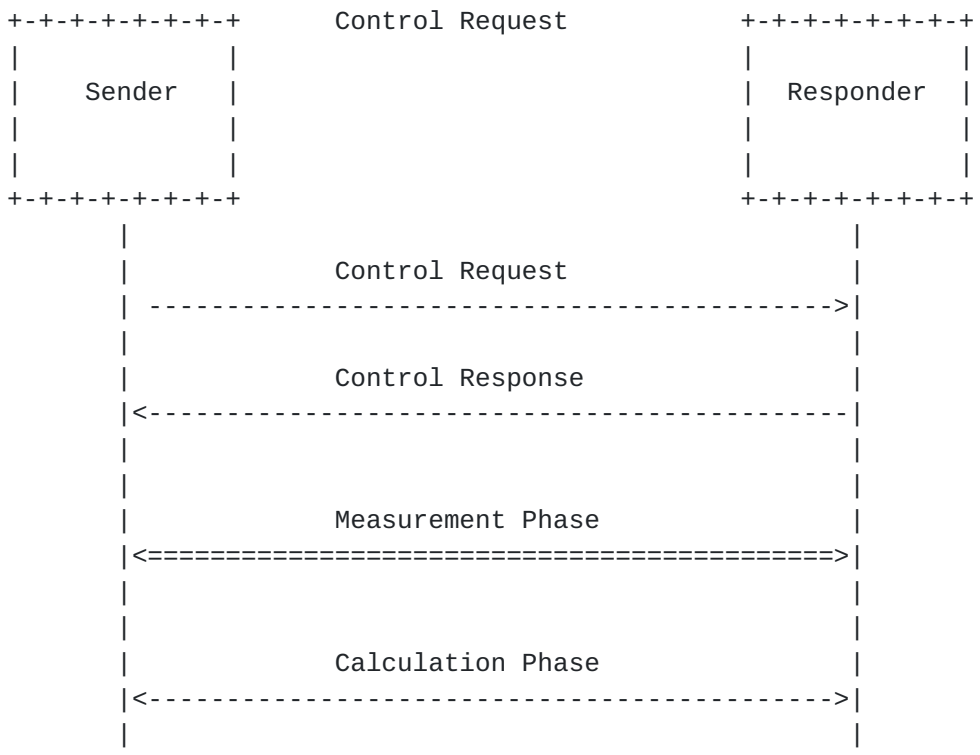
- Local assessment: Sender initiates a Calculation Request message and Responder sends back a Calculation Response message. Sender and Responder, upon receipt test traffic, create data structure with timestamped records then computes service level metrics from that data structure. Let's call this data structure the test receipt.

- Central assessment: A "central" entity (e.g. a controller) compares the test receipt collected by the Responder with data structure obtained from the Sender, then computes the service levels by means of comparing.

- Local assessment with reference recording: Both sender and receiver play out the same test traffic. Assessment is done locally not by computing metrics over the test receipt, but by "overlaying" the original with the one that was received and computing the delta.

The number and frequency with which messages are sent SHOULD be controlled by configuration on the Sender element, along with the waiting time for a Control Response.

The following sequence diagram depicts the message exchanges:



To utilize Cisco SLA Protocol, some extensions are needed. As part of the Control Request, the Sender needs to indicate that it will send test traffic to be analyzed. However, it indicates that alternate marking techniques are to be used and that traffic is going to be marked. Likewise, it can indicate to the Responder to not simply reflect the marked traffic, but to generate a separate stream of marked test traffic back to the sender. The marking pattern will be conveyed (including the alternate markings to be used and duration of the marking intervals). The implementation of measurements involves analyzing the marked traffic as needed. Conveying of results of the analysis of observed traffic occurs through separate means, not specified here.

4.1. Control Phase

The Control Phase, as described in RFC 6812 [RFC6812] section 3.1, begins with the Sender sending a Control-Request message to the Responder. The Responder replies by sending a Control-Response with an appropriate Status indicating Success when the Sender identity is verified and the requested UDP port was successfully opened. In all other cases, a non-zero Status is returned in the Command-Header Status field.

4.1.1. Control Request

The Control Request message, as described in RFC 6812 [RFC6812] section 3.1.1, consists of a Command Header followed by one or more Command, Status, Length and Data sections (henceforth known as CSLD). At the minimum, there SHOULD be at the least two CSLD sections, one of which is the authentication CSLD section and the other carries information for the Measurement Phase simulation type.

4.1.1.1. Command Header

The Command Header is the first section of the Control Request message and is described in RFC 6812 [RFC6812] section 3.1.1.1.

4.1.1.2. CSLDs

The two CSLDs to be included, in order, along with the Command Header are:

- The Authentication CSLD
- A Measurement Type CSLD

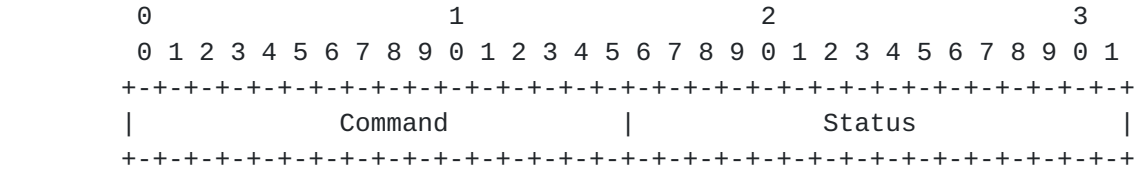
In this revision of the protocol, an additional Measurement Type CSLD has been defined, the Alternate Marking Measurement Type CSLD. In RFC 6812 [RFC6812] section 3.1.1.2 is described the UDP-Measurement CSLD.

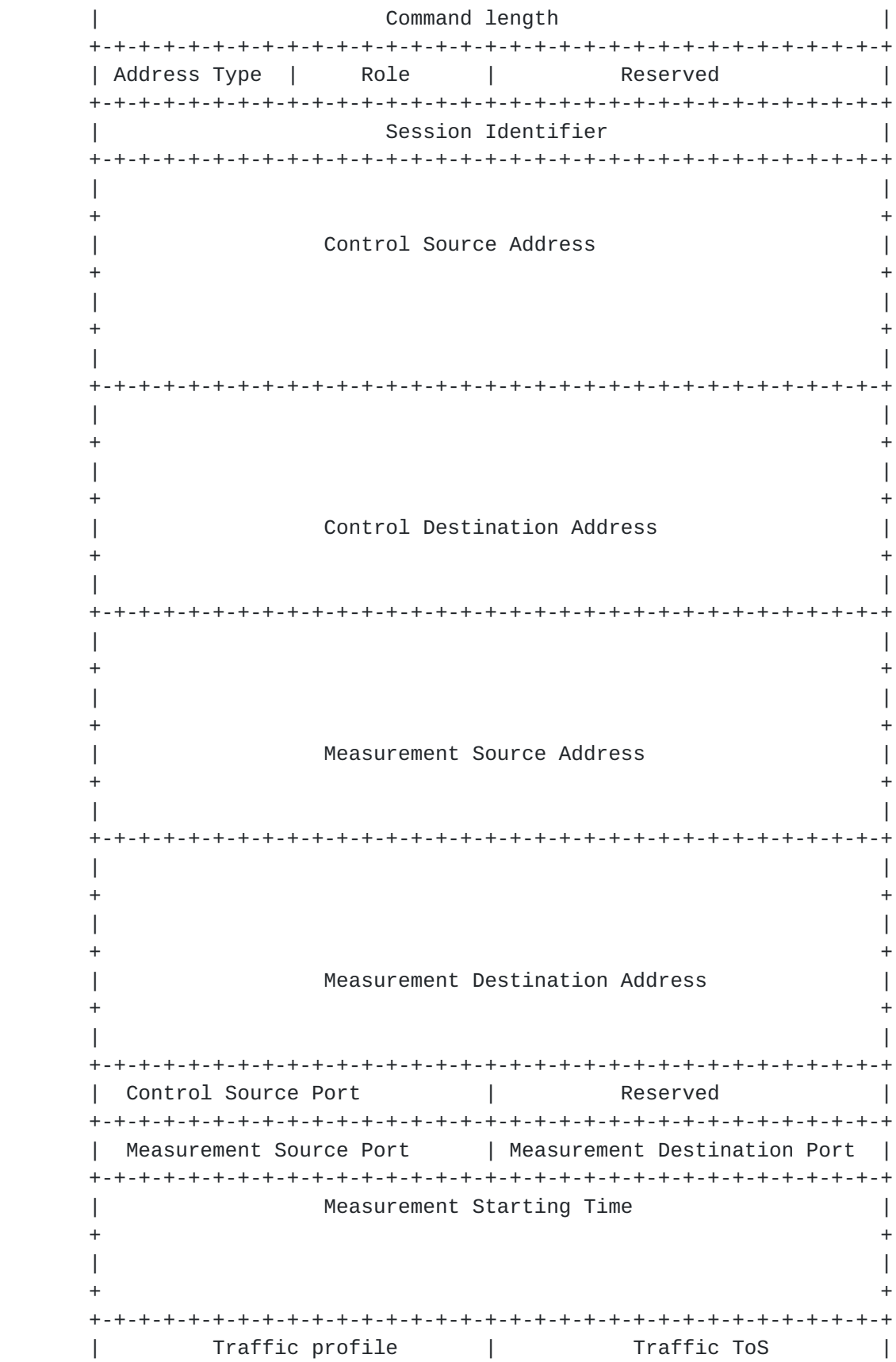
4.1.1.2.1. Authentication CSLD

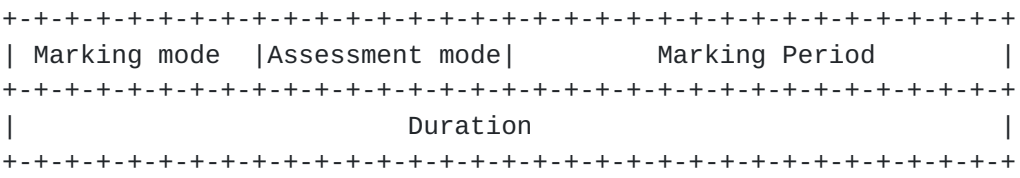
The Authentication CSLD provides the message authentication and verifies the requester knows the shared-secret. The format for the Authentication CSLD is detailed in RFC 6812 [RFC6812] section 3.1.1.2.1.

4.1.1.2.2. Alternate Marking Measurement CSLD

The Alternate Marking Measurement CSLD indicates the Measurement Type to be used during the Measurement Phase and specifies the addresses and UDP port to be opened as well as the duration the port has to be kept open for the measurement phase. The format of the CSLD is as follows:




Command length
Address Type
Role
Reserved
Session Identifier
Control Source Address
Control Destination Address
Measurement Source Address
Measurement Destination Address
Control Source Port
Reserved
Measurement Source Port
Measurement Destination Port
Measurement Starting Time
Traffic profile
Traffic ToS



The new fields that have been added to RFC 6812 [RFC6812] section 3.1.1.2.2 are: Measurement Starting time, Traffic Profile, Traffic TOS, Marking mode, Assesment mode, Marking period. They are described below.

The unchanged fields are detailed in RFC 6812 [RFC6812] section 3.1.1.2.2:

So the new fields in the Alternate Marking Measurement CSLD have the following meaning:

Field	Size (bits)	Definition
Command	16	Indicates that the CSLD is to simulate UDP alternate marking traffic measurements.
---	---	---
Measurement Starting Time	64	Carries the timestamp when the Measurement Phase will start
---	---	---
Traffic Profile	16	Indicates a fixed profile with an assigned value (defined outside this document), to establish size, number of packets, milliseconds between packets that are generated in both directions.
---	---	---
Traffic ToS	16	Indicates the Type of Service of the generated test frames: but if the marking field is the ToS field, the two marking ToS values are the first and the last 8 bits; otherwise if the marking field is different, the first 8 bits are zero and the last 8 bits indicates the ToS of all the generated frames.
---	---	---
Marking mode	8	Indicates one of the alternatives for Marking Field: marking IPv4 header (Type of Service Field or the last reserved bit of the Flag field) or marking UDP payload (Measurement Type or Data).
---	---	---
Assessment mode	8	Indicates one of the three alternatives for the Calculation Phase: 1 - Local assessment, 2 - Central assessment, 3 - Local assessment with reference recording.
---	---	---
Marking Period	16	Indicates the duration in seconds of the Alternate Marking period

Note: The source addresses are only indicative of identity of the originator and cannot be used as destination for responses in a NAT environment.

Note: In case of Local assessment with reference recording, Sender and Responder exchanges the refernce reconding before the Measurement Phase.

Note: All timestamps have the format as described in RFC 5905 [RFC5905] and is as follows: the first 32 bits represent the unsigned integer number of seconds elapsed since 0h on 1 January 1900; the next 32 bits represent the fractional part of a second thereof. The timestamp definition is also similar to RFC 4656 [RFC4656]

In addition, the timestamp format used can be as described for the low-order 64 bits of the IEEE 1588-2008 (1588v2) Precision Time Protocol timestamp format [IEEE1588]. This truncated format consists of a 32-bit seconds field followed by a 32-bit nanoseconds field, and is the same as the IEEE 1588v1 timestamp format. This timestamp definition is similar to the default timestamp as specified in RFC 6374 [RFC6374]

Implementations MUST use only one of the two formats. The chosen format is negotiated out-of-band between the endpoints.

4.1.2. Control Response Message

In response to the Control Request Message the network element designated the Responder sends back a Control Response Message that reflects the Command Header with an updated Status field and includes the two CSLD sections that also carry updated Status fields. Hence, the format is identical to the Control Request message as described above. The supported values of the Status fields are the same described in RFC 6812 [RFC6812] section 3.1.2.

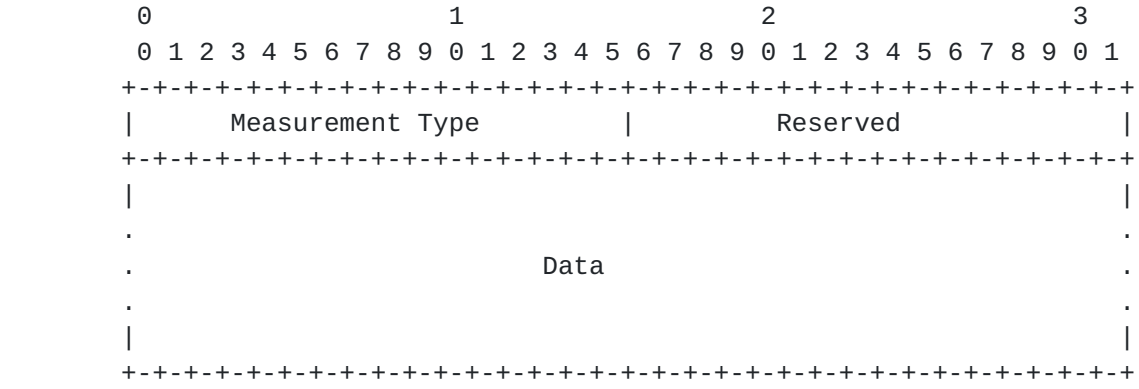
4.2. Measurement Phase

Upon receiving the Control Response message with the Status set to Success, the second phase of the protocol, the Measurement Phase, is initiated. In all other cases when the Status is not success no measurement traffic is initiated. In the Measurement Phase the Sender sends a stream of measurement messages. The measurement message stream consists of packets/frames that are spaced a configured number of milliseconds.

The Measurement messages as defined by this document for Alternate Marking UDP Measurements is as shown below and is simplified in comparison to RFC 6812 [RFC6812] section 3.2. In particular the fields that have been removed from RFC 6812 [RFC6812] section 3.2 are: Sender Send Time, Responder Receive Time, Responder Send Time, Sender Receive Time, Sender Clock Offset, Responder Clock Offset, Sender Sequence No. and Responder Sequence No.

The format of the Measurement messages is the same for the exchange in both directions, that is when sent from the Sender to the Responder and from the Responder to the Sender.

Note: Marking field can be chosen in two ways: marking UDP payload or marking IPv4 header. Marking IPv4 header (Type of Service Field or the last reserved bit of the Flag field) is useful so in this way the active measurement could use the same functions of passive measurement.



The fields for the UDP Measurement message have the following meaning:

Field	Size (bits)	Description
Measurement Type	16	Carries the type of measurement being performed (This field can include Marking: at least two values); 1 - Reserved, 2 - Reserved, 3 - UDP
---------	---------	-------------------------
Reserved	16	Reserved field and MUST be set to 0
---------	---------	-------------------------
Data	32 bit aligned	This field is used to pad up to the configured request data size. The minimum requested data size SHOULD be 512 bytes and this field will be of length 512 minus the length of the previous fields. This field can include Marking

Note: No timestamp, No sequence number. The two data flows are indipendent.

4.3. Calculation Phase

As mentioned above, the Calculation Phase is introduced "ad hoc" for Alternate Marking implementation because it does not exist in Cisco Service Level Assurance Protocol described in RFC 6812 [RFC6812]. After test execution there are some alternatives to compute packet loss, delay and delay variation:

- Local assessment: Sender initiates a Calculation Request message and Responder sends back a Calculation Response message. Sender and Responder, upon receipt test traffic, create data structure with timestamped records then computes service level metrics from that data structure. Let's call this data structure the test receipt).

- Central assessment: A "central" entity (e.g. a controller) compares the test receipt collected by the Responder with data structure obtained from the Sender, then computes the service levels by means of comparing.

- Local assessment with reference recording: Both sender and receiver play out the same test traffic. Assessment is done locally not by computing metrics over the test receipt, but by "overlaying" the original with the one that was received and computing the delta.

5. Implementation notes

Implementation notes are detailed in RFC 6812 [RFC6812] section 4.

6. IANA Considerations

IANA needs to reserve a new value for Alternate Marking CSLD Command Registry. The available values for future extensions are detailed in RFC 6812 [RFC6812] section 6.

7. Security Considerations

Security Considerations are detailed in RFC 6812 [RFC6812] section 7.

8. Terminology

Term	Description
Control Phase	A phase during which Control Request and Control Response is exchanged.
L2	OSI Data Link Layer
L3	OSI Network Layer
Measurement Phase	Active measurement phase that is marked by a sequence of Measurement Request and Measurement Response exchanges.
Metric	A particular characteristic of the network data traffic, for example latency, jitter, packet/frame loss
Responder	A network element that responds to a message
RTP	Real-time Transport Protocol
Sender	A network element that is the initiator of a message exchange
Service Level	This is the level of service that is agreed upon between the Provider and the Customer
UDP	User Datagram Protocol

9. Acknowledgements

Thanks to Luca Castaldelli, Francesco Burgio and Stefano Righetti from Telecom Italia for their contribution to the prototype implementation of the method.

Mauro Cociglio and Giuseppe Fioccola worked in part on the Leone research project, which received funding from the European Union Seventh Framework Programme [FP7/2007-2013] under grant agreement number 317647.

10. References

10.1. Normative References

[I-D.tempia-ippm-p3m]
Capello, A., Cociglio, M., Fioccola, G., Castaldelli, L., and A. Bonda, "A packet based method for passive performance monitoring", draft-tempia-ippm-p3m-02 (work in progress), October 2015.

[IEEE1588]
IEEE, "1588-2008 Standard for a Precision Clock Synchronization Protocol for Networked Measurement and Control Systems", March 2008.

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

[RFC6812] Chiba, M., Clemm, A., Medley, S., Salowey, J., Thombare, S., and E. Yedavalli, "Cisco Service-Level Assurance Protocol", RFC 6812, DOI 10.17487/RFC6812, January 2013, <http://www.rfc-editor.org/info/rfc6812>.

10.2. Informative References

[RFC3579] Aboba, B. and P. Calhoun, "RADIUS (Remote Authentication Dial In User Service) Support For Extensible Authentication Protocol (EAP)", RFC 3579, DOI 10.17487/RFC3579, September 2003, <http://www.rfc-editor.org/info/rfc3579>.

[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, <http://www.rfc-editor.org/info/rfc4656>.

[RFC4868] Kelly, S. and S. Frankel, "Using HMAC-SHA-256, HMAC-SHA-384, and HMAC-SHA-512 with IPsec", RFC 4868, DOI 10.17487/RFC4868, May 2007, <http://www.rfc-editor.org/info/rfc4868>.

[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, <http://www.rfc-editor.org/info/rfc5357>.

[RFC5905] Mills, D., Martin, J., Ed., Burbank, J., and W. Kasch, "Network Time Protocol Version 4: Protocol and Algorithms Specification", RFC 5905, DOI 10.17487/RFC5905, June 2010, <http://www.rfc-editor.org/info/rfc5905>.

[RFC6374] Frost, D. and S. Bryant, "Packet Loss and Delay Measurement for MPLS Networks", RFC 6374, DOI 10.17487/RFC6374, September 2011, <http://www.rfc-editor.org/info/rfc6374>.

[RFC6703] Morton, A., Ramachandran, G., and G. Maguluri, "Reporting IP Network Performance Metrics: Different Points of View", RFC 6703, DOI 10.17487/RFC6703, August 2012, <http://www.rfc-editor.org/info/rfc6703>.

Authors' Addresses

Giuseppe Fioccola
Telecom Italia
Via Reiss Romoli, 274
Torino 10148
Italy

Email: giuseppe.fioccola@telecomitalia.it

Alexander Clemm
Cisco Systems
170 West Tasman Drive
San Jose 95134
USA

Phone: 1-408-526-4000
Email: alex@cisco.com

Mauro Cociglio
Telecom Italia
Via Reiss Romoli, 274
Torino 10148
Italy

Email: mauro.cociglio@telecomitalia.it

Mouli Chandramouli
Cisco Systems

Email: moulchan@cisco.com

Alessandro Capello
Telecom Italia
Via Reiss Romoli, 274
Torino 10148
Italy

Email: alessandro.capello@telecomitalia.it