Network Working Group
Internet-Draft
Intended status: Informational
Expires: July 25, 2014

P. Eardley
BT
A. Morton
AT&T Labs
M. Bagnulo
UC3M
T. Burbridge
BT
P. Aitken
A. Akhter
Cisco Systems

January 21, 2014

A framework for large-scale measurement platforms (LMAP)
draft-ietf-lmap-framework-03

Abstract

Measuring broadband service on a large scale requires a description of the logical architecture and standardisation of the key protocols that coordinate interactions between the components. The document presents an overall framework for large-scale measurements. It also defines terminology for LMAP (large-scale measurement platforms).

Table of Contents

1. Introduction

There is a desire to be able to coordinate the execution of broadband measurements and the collection of measurement results across a large scale set of diverse devices. These devices could be software based agents on PCs, embedded agents in consumer devices (e.g. blu-ray players), service provider controlled devices such as set-top players and home gateways, or simply dedicated probes. It is expected that such a system could easily comprise 100k devices. Such a scale presents unique problems in coordination, execution and measurement result collection. Several use cases have been proposed for large-scale measurements including:

- Operators: to help plan their network and identify faults

- Regulators: to benchmark several network operators and support public policy development

Further details of the use cases can be found at [I-D.ietf-lmap-use-cases]. The LMAP framework should be useful for these, as well as other use cases that the LMAP WG doesn't concentrate on, such as to help end users run diagnostic checks like a network speed test.

The LMAP framework has four basic elements: Measurement Agents, Measurement Peers, Controllers and Collectors.

Measurement Agents (MAs) perform Measurement Tasks, perhaps in conjunction with Measurement Peers. They are pieces of code that can be executed in specialized hardware (hardware probe) or on a general-purpose device (like a PC or mobile phone). A device with a Measurement Agent may have multiple interfaces (WiFi, Ethernet, DSL, fibre, etc.) and the Measurement Tasks may specify any one of these. Measurement Tasks may be Active (the MA or Measurement Peer generates Active Measurement Traffic), Passive (the MA observes user traffic), or some hybrid form of the two. For Active Measurement Tasks, the MA (or Measurement Peer) generates Active Measurement Traffic and measures some metric associated with its transfer over the path to (or from) a Measurement Peer. For example, one Active Measurement Task could be to measure the UDP latency between the MA and a given Measurement Peer. MAs may also conduct Passive Measurement Tasks through the observation of traffic. The Measurement Tasks themselves may be on IPv4, IPv6, and on various services (DNS, HTTP, XMPP, FTP, VoIP, etc.).

The Controller manages one or more MAs by instructing it which Measurement Tasks it should perform and when. For example it may instruct a MA at a home gateway: "Measure the 'UDP latency' with the Measurement Peer mp.example.org; repeat every hour at xx.05". The Controller also manages a MA by instructing it how to report the Measurement Results, for example: "Report results once a day in a batch at 4am". We refer to these as the Measurement Schedule and Report Schedule.

The Collector accepts Reports from the MAs with the Results from their Measurement Tasks. Therefore the MA is a device that gets Instructions from the Controller, initiates the Measurement Tasks, and reports to the Collector.

There are additional elements that are part of a measurement system, but that are out of the scope for LMAP. We provide a detailed discussion of all the elements in the rest of the document.

The desirable features for a large-scale measurement systems we are designing for are:

- Standardised - in terms of the Measurement Tasks that they perform, the components, the data models and protocols for transferring information between the components. Amongst other things, standardisation enables meaningful comparisons of measurements made of the same metric at different times and places, and enables the operator of a measurement system to buy the various components from different vendors. Today's systems are proprietary in some or all of these aspects.

- Large-scale - [I-D.ietf-lmap-use-cases] envisages Measurement Agents in every home gateway and edge device such as set-top-boxes and tablet computers. It is expected that a measurement system could easily encompass a few hundred thousand Measurement Agents. Existing systems have up to a few thousand MAs (without judging how much further they could scale).

- Diversity - a measurement system should handle different types of Measurement Agent - for example Measurement Agents may come from different vendors, be in wired and wireless networks and be on devices with IPv4 or IPv6 addresses.

2. Outline of an LMAP-based measurement system

Figure 1 shows the main components of a measurement system, and the interactions of those components. Some of the components are outside the scope of LMAP. In this section we provide an overview of the whole measurement system and we introduce the main terms needed for the LMAP framework. The new terms are capitalized. In the next section we provide a terminology section with a compilation of all the LMAP terms and their definition. The subsequent sections study the LMAP components in more detail.

A Measurement Task measures some performance or reliability Metric of interest. An Active Measurement Task involves either a Measurement Agent (MA) injecting Active Measurement Traffic into the network destined for a Measurement Peer, and/or a Measurement Peer sending Active Measurement Traffic to a MA; one of them measures some parameter associated with the transfer of the packet(s). A Passive Measurement Task involves only a MA, which simply observes existing traffic - for example, it could simply count bytes or it might calculate the average loss for a particular flow.

It is very useful to standardise Measurement Methods (a Measurement Method is a generalisation of a Measurement Task), so that it is meaningful to compare measurements of the same Metric made at different times and places. It is also useful to define a registry for commonly-used Metrics [I-D.bagnulo-ippm-new-registry-independent] so that a Measurement Method can be referred to simply by its identifier in the registry. The Measurement Methods and registry will hopefully be referenced by other standards organisations.

In order for a Measurement Agent and a Measurement Peer to execute an Active Measurement Task, they exchange Active Measurement Traffic. The protocols used for the Active Measurement Traffic is out of the scope of the LMAP WG and falls within the scope of other IETF WGs such as IPPM.

For Measurement Results to be truly comparable, as might be required by a regulator, not only do the same Measurement Methods need to be used but also the set of Measurement Tasks should follow a similar Measurement Schedule and be of similar number. The details of such a characterisation plan are beyond the scope of work in IETF although certainly facilitated by IETF's work.

The next components we consider are the Measurement Agent (MA), Controller and Collector. The main work of the LMAP working group is to define the Control Protocol between the Controller and MA, and the Report Protocol between the MA and Collector. Section 4 onwards considers the LMAP components in more detail; here we introduce them.

The Controller manages a MA by instructing it which Measurement Tasks it should perform and when. For example it may instruct a MA at a home gateway: "Run the 'download speed test' with the Measurement Peer at the end user's first IP point in the network; if the end user is active then delay the test and re-try 1 minute later, with up to 3 re-tries; repeat every hour at xx.05 + Unif[0,180] seconds". The Controller also manages a MA by instructing it how to report the Measurement Results, for example: "Report results once a day in a batch at 4am + Unif[0,180] seconds; if the end user is active then delay the report 5 minutes". These are called the Measurement and Report Schedule. As well as periodic Measurement Tasks, a Controller can initiate a one-off (non-recurring) Measurement Task ("Do measurement now", "Report as soon as possible").

The Collector accepts a Report from a MA with the results from its Measurement Tasks. It may also do some post-processing on the results, for instance to eliminate outliers, as they can severely impact the aggregated results.

Finally we introduce several components that are out of scope of the LMAP WG and will be provided through existing protocols or applications. They affect how the measurement system uses the Measurement Results and how it decides what set of Measurement Tasks to perform.

The MA needs to be bootstrapped with initial details about its Controller, including authentication credentials. The LMAP WG considers the bootstrap process, since it affects the Information Model. However, it does not define a bootstrap protocol, since it is likely to be technology specific and could be defined by the Broadband Forum, CableLabs or IEEE depending on the device. Possible protocols are SNMP, NETCONF or (for Home Gateways) CPE WAN Management Protocol (CWMP) from the Auto Configuration Server (ACS) (as specified in TR-069).

A Subscriber parameter database contains information about the line, such as the customer's broadband contract (perhaps 2, 40 or 80Mb/s), the line technology (DSL or fibre), the time zone where the MA is located, and the type of home gateway and MA. These parameters are already gathered and stored by existing operations systems. They may affect the choice of what Measurement Tasks to run and how to interpret the Measurement Results. For example, a download test suitable for a line with an 80Mb/s contract may overwhelm a 2Mb/s line.

A results repository records all Measurement Results in an equivalent form, for example an SQL database, so that they can easily be accessed by the data analysis tools. The data analysis tools also need to understand the Subscriber's service information, for example the broadband contract.

The data analysis tools receive the results from the Collector or via the Results repository. They might visualise the data or identify which component or link is likely to be the cause of a fault or degradation. This information could help the Controller decide what follow-up Measurement Task to perform in order to diagnose a fault.

The operator's OAM (Operations, Administration, and Maintenance) uses the results from the tools.

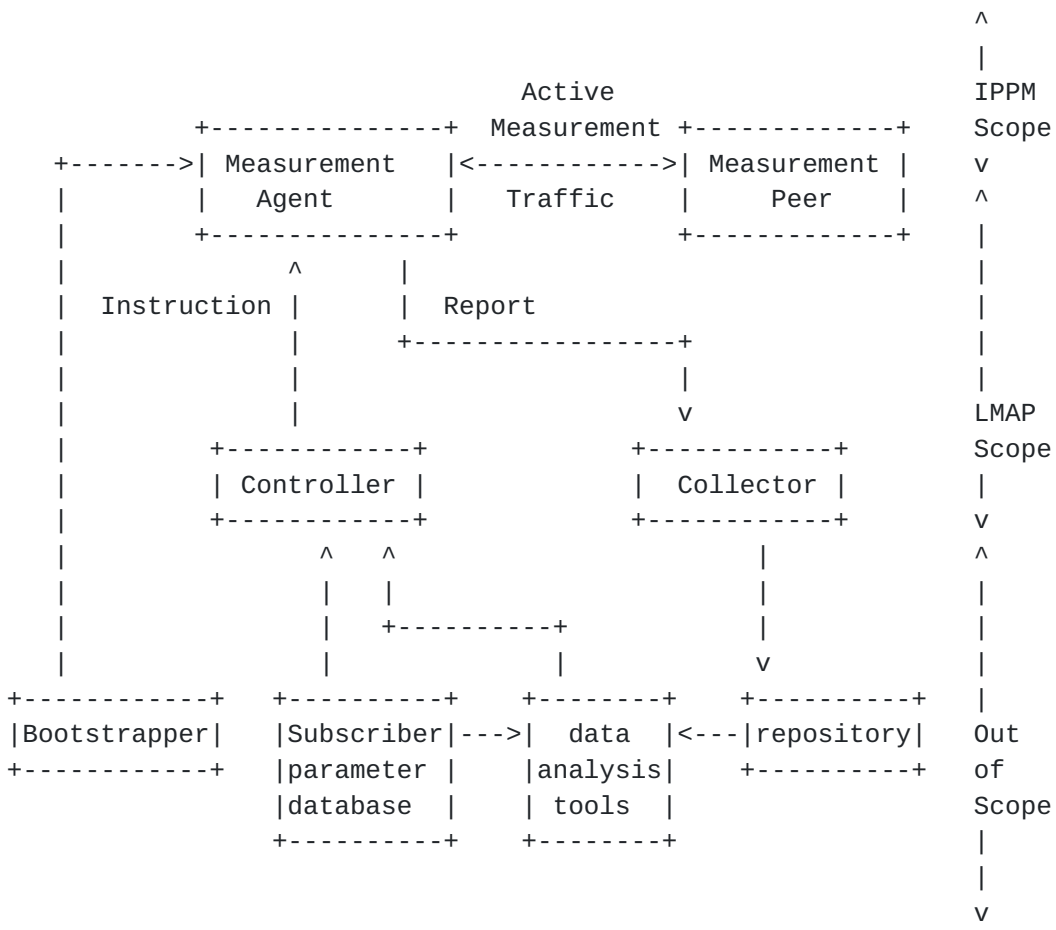


Figure 1: Schematic of main elements of an LMAP-based measurement system
(showing the elements in and out of the scope of the LMAP WG)

3. Terminology

This section defines terminology for LMAP. Please note that defined terms are capitalized.

Active Measurement Method (Task): A type of Measurement Method (Task) that involves a Measurement Agent and a Measurement Peer (or possibly Peers), where either the Measurement Agent or the Measurement Peer injects Active Measurement Traffic into the network destined for the other, and which involves one of them measuring some performance or reliability parameter associated with the transfer of the traffic.

Active Measurement Traffic: the packet(s) generated by the Measurement Agent and/or the Measurement Peer, as part of an Active Measurement Task.

Bootstrap: A process that initialises a Measurement Agent with the information necessary to be integrated into a measurement system.

Capabilities: Information about the Measurement Methods that the MA can perform and the device hosting the MA, for example its interface type and speed and its IP address.

Channel: an Instruction Channel, Report Channel or MA-to-Controller Channel

Collector: A function that receives a Report from a Measurement Agent.

Composite Metric: A Metric that is a combination of other Metrics, and/or a combination of the same Metric measured over different parts of the network or at different times.

Controller: A function that provides a Measurement Agent with Instruction(s).

Control Channel: a communications channel between a Controller and a MA, which is defined by a specific Controller, MA and associated security, and over which Instructions are sent.

Control Protocol: The protocol delivering Instruction(s) from a Controller to a Measurement Agent. It also delivers Failure Information and Capabilities Information from the Measurement Agent to the Controller.

Cycle-ID: (optional) A tag that is sent by the Controller in an Instruction and echoed by the MA in its Report. The same Cycle-ID is used by several MAs that use the same Measurement Method with the same Input Parameters. Hence the Cycle-ID allows the Collector to easily identify Measurement Results that should be comparable.

Data Model: The implementation of an Information Model in a particular data modelling language.

Environmental Constraint: A parameter that is measured as part of the Measurement Task, its value determining whether the rest of the Measurement Task proceeds.

Failure Information: Information about the MA's failure to action or execute an Instruction, whether concerning Measurement Tasks or Reporting.

Group-ID: (optional) An identifier of a group of MAs.

Information Model: The protocol-neutral definition of the semantics of the Instructions, the Report, the status of the different elements of the measurement system as well of the events in the system.

Input Parameter: A parameter whose value is left open by the Measurement Method and is set to a specific value in a Measurement Task. Altering the value of an Input Parameter does not change the fundamental nature of the Measurement Method.

Instruction: The description of Measurement Tasks to perform and the details of the Report to send. The Instruction is sent by a Controller to a Measurement Agent.

MA-to-Controller Channel: a communications channel between a MA and a Controller, which is defined by a specific Controller, MA and associated security, and over which Capabilities and Failure Information is sent.

Measurement Agent (MA): The function that receives Instructions from a Controller, performs Measurement Tasks (perhaps in concert with a Measurement Peer) and reports Measurement Results to a Collector.

Measurement Agent Identifier (MA-ID): a UUID [RFC4122], which is configured as part of the Bootstrapping and included in a Capabilities message, Failure Information message and optionally in a Report.

Measurement Method: The process for assessing the value of a Metric; the process of measuring some performance or reliability parameter; the generalisation of a Measurement Task.

Measurement Peer: The function that receives control messages and Active Measurement Traffic from a Measurement Agent and may reply to the Measurement Agent as defined by the Active Measurement Method.

Measurement Result: The output of a single Measurement Task (the value obtained for the parameter of interest or Metric).

Measurement Schedule: the schedule for performing Measurement Tasks.

Measurement Suppression: a type of Instruction that temporarily stops (suppresses) Active Measurement Tasks.

Measurement Task: The act that yields a single Measurement Result; the act consisting of the (single) operation of the Measurement Method at a particular time and with all its parameters set to specific values.

Metric: The quantity related to the performance and reliability of the network that we'd like to know the value of, and that is carefully specified.

Passive Measurement Method (Task): A Measurement Method (Task) in which a Measurement Agent observes existing traffic but does not inject Active Measurement Traffic.

Report: The Measurement Results and other associated information (as defined by the Instruction). The Report is sent by a Measurement Agent to a Collector.

Report Channel: a communications channel between a MA and a Collector, which is defined by a specific MA, Collector, Report Schedule and associated security, and over which Reports are sent.

Report Protocol: The protocol delivering Report(s) from a Measurement Agent to a Collector.

Report Schedule: the schedule for sending one or more Reports to a Collector.

Subscriber: An entity (associated with one or more users) that is engaged in a subscription with a service provider. The Subscriber is allowed to subscribe and un-subscribe services, and to register a user or a list of users authorized to enjoy these services. [Q1741] Both the Subscriber and service provider are allowed to set the limits relative to the use that associated users make of subscribed services.

4. Constraints

The LMAP framework makes some important assumptions, which constrain the scope of the work to be done.

4.1. Measurement system is under the direction of a single organisation

In the LMAP framework, the measurement system is under the direction of a single organisation that is responsible both for the data and the quality of experience delivered to its users. Clear responsibility is critical given that a misbehaving large-scale measurement system could potentially harm user experience, user privacy and network security.

However, the components of an LMAP measurement system can be deployed in administrative domains that are not owned by the measuring organisation. Thus, the system of functions deployed by a single

organisation constitutes a single LMAP domain which may span ownership or other administrative boundaries.

4.2. Each MA may only have a single Controller at any point in time

A MA is instructed by one Controller and is in one measurement system. The constraint avoids different Controllers giving a MA conflicting instructions and so means that the MA does not have to manage contention between multiple Measurement (or Report) Schedules. This simplifies the design of MAs (critical for a large-scale infrastructure) and allows a Measurement Schedule to be tested on specific types of MA before deployment to ensure that the end user experience is not impacted (due to CPU, memory or broadband-product constraints).

An operator may have several Controllers, perhaps with a Controller for different types of MA (home gateways, tablets) or location (Ipswich, Edinburgh).

5. LMAP Protocol Model

A protocol model [RFC4101] presents an architectural model for how the protocol operates and needs to answer three basic questions:

1. What problem is the protocol trying to achieve?

2. What messages are being transmitted and what do they mean?

3. What are the important, but unobvious, features of the protocol?

An LMAP system goes through the following phases:

- a bootstrapping process before the MA can take part in the other three phases

- a Control Protocol, which delivers an Instruction from a Controller to a MA, detailing what Measurement Tasks the MA should perform and when, and how it should report the Measurement Results

- the actual Measurement Tasks are performed. An Active Measurement Task involves sending Active Measurement Traffic between the Measurement Agent and a Measurement Peer, whilst a Passive Measurement Task involves (only) the Measurement Agent observing existing user traffic. The LMAP WG does not define Measurement Methods, however the IPPM WG does.

- a Report Protocol, which delivers a Report from the MA to a Collector. The Report contains the Measurement Results.

In the diagrams the following convention is used:

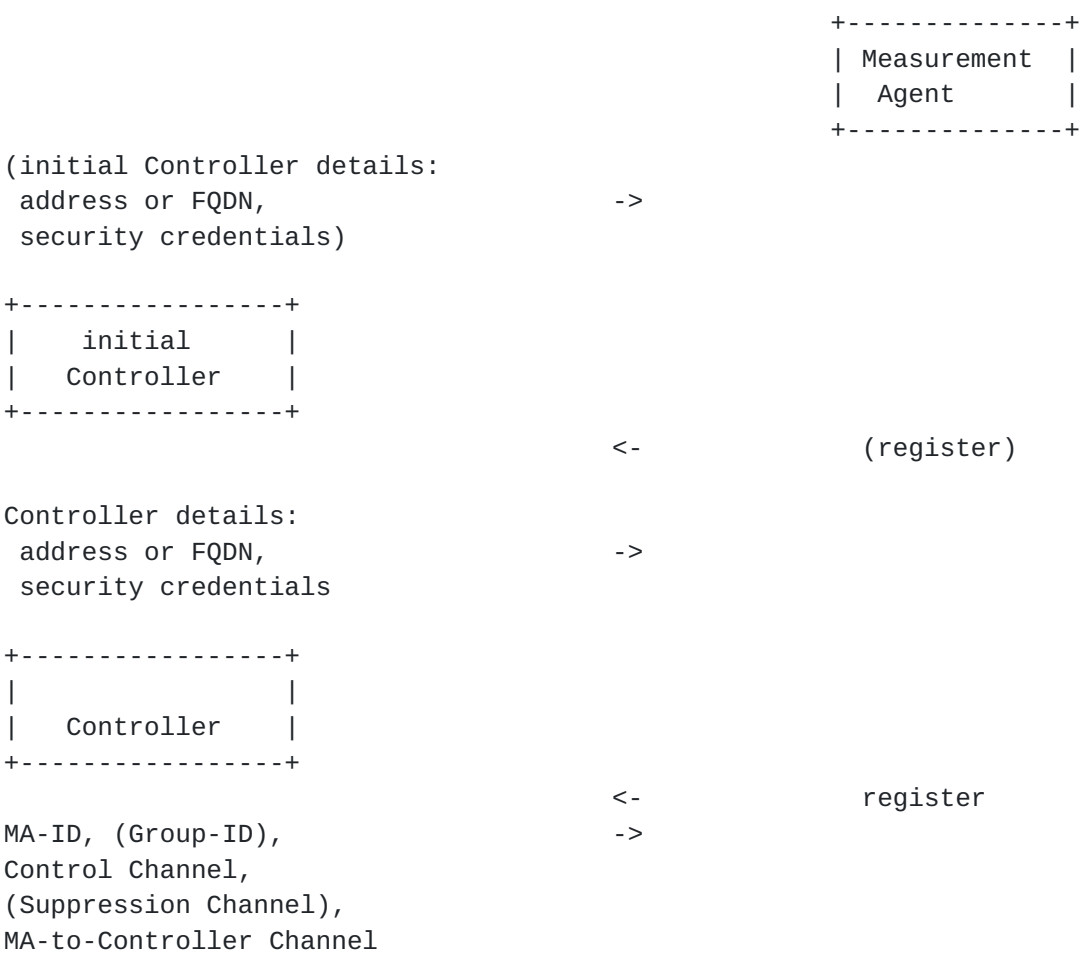
- (optional): indicated by round brackets
- [potentially repeated]: indicated by square brackets

The protocol model is closely related to the Information Model [I-D.burbridge-lmap-information-model], which is the abstract definition of the information carried by the protocol model. The purpose of both is to provide a protocol and device independent view, which can be implemented via specific protocols. The LMAP WG will define a specific Control Protocol and Report Protocol, but others could be defined by other standards bodies or be proprietary. However it is important that they all implement the same Information Model and protocol model, in order to ease the definition, operation and interoperability of large-scale measurement systems.

The diagrams show the various LMAP messages and Section 5.5 considers how they could be mapped onto an underlying transport protocol.

5.1. Bootstrapping process

The primary purpose of bootstrapping is to enable the MA and Controller to be integrated into a measurement system. In order to do that, the MA needs to retrieve information about itself (like its identity in the measurement system), about the Controller, as well as security information (such as certificates and credentials).



The MA knows how to contact a Controller through some device /access specific mechanism. For example, this could be in the firmware, downloaded, manually configured or via a protocol like TR-069. The Controller could either be the one that will send it Instructions or else an initial Controller (whose details may be statically configured). The role of an initial Controller is simply to inform the MA how to contact its actual Controller, for example its FQDN (Fully Qualified Domain Name) [RFC1035].

The MA learns its identifier (MA-ID). It may also be told a Group-ID and whether to include the MA-ID as well as the Group-ID in its Reports. A Group-ID would be shared by several MAs and could be useful for privacy reasons, for instance to hinder tracking of a mobile device.

The MA is also told about the Control Channel over which it will receive Instructions from the Controller, in particular the associated security information, for example to enable the MA to decrypt the Instruction. Optionally any Suppression messages can be sent over a different Channel. The MA is also informed about the MA-

to-Controller Channel, over which the MA can tell the Controller about its Capabilities and any Failure Information. This consists of the address of the Controller, for instance its URL, and security details for MA-to-Controller messages.

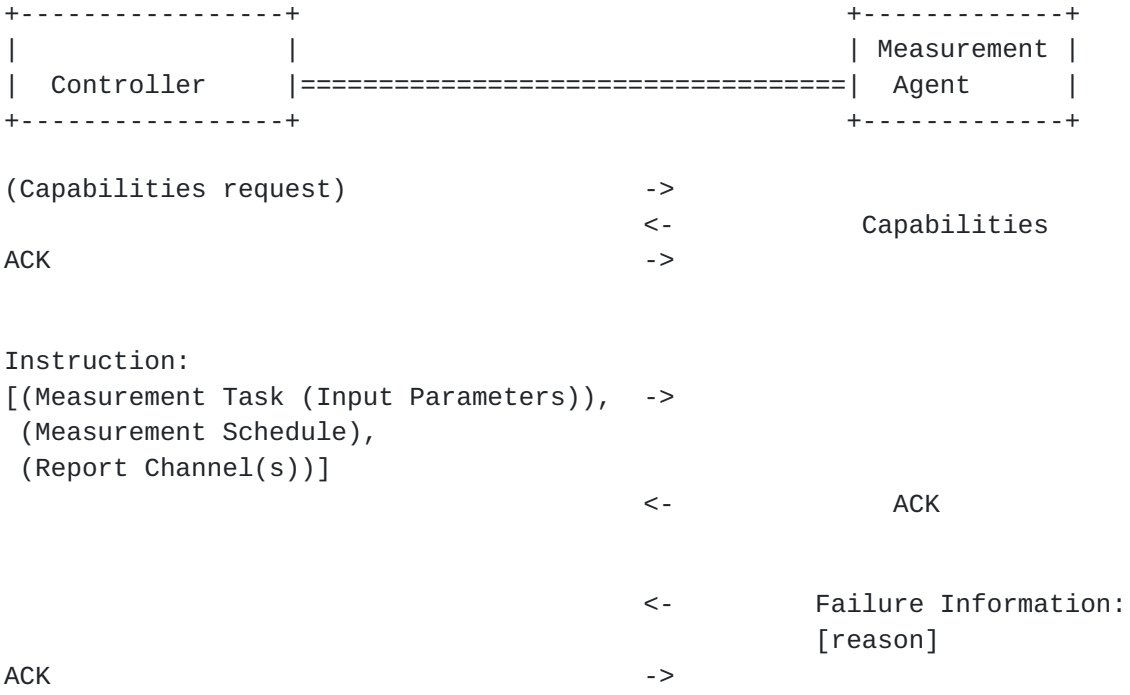
The MA may tell the Controller its Capabilities, in particular the Measurement Methods it can perform.

If the device with the MA re-boots, then the MA needs to re-register, so that it can receive a new Instruction. To avoid a "mass calling event" after a widespread power restoration affecting many MAs, it is sensible for an MA to pause for a random delay (perhaps in the range of one minute or so) before re-registering.

Whilst the LMAP WG considers the bootstrapping process, it is out of scope to define a bootstrap mechanism, as it depends on the type of device and access.

5.2. Control Protocol

The primary purpose of the Control Protocol is to allow the Controller to configure a Measurement Agent with an Instruction about what Measurement Tasks to do, when to do them, and how to report the Measurement Results. The Measurement Agent then acts on the Instruction autonomously.



The Controller needs to know the Capabilities of the MA, and in particular what Measurement Methods it supports, so that it can correctly instruct the MA. It is possible that the Controller knows the MA's Capabilities via some mechanism beyond the scope of LMAP, such as a device-specific protocol. In LMAP, the MA can inform the Controller about its Capabilities. This message could be sent in several circumstances: when the MA first communicates with a Controller; when the MA becomes capable of a new Measurement Method; when requested by the Controller (for example, if the Controller forgets what the MA can do or otherwise wants to resynchronize what it knows about the MA). Note that Capabilities do not include dynamic information like the MA's currently unused CPU, memory or battery life.

A single Instruction message contains one, two, three or all four of the following elements:

- configuration of all the Measurement Tasks, each of which needs:
  - the Measurement Method, specified as a URN to a registry entry. The registry could be defined by the IETF [I-D.bagnulo-ippm-new-registry-independent], locally by the operator of the measurement system or perhaps by another standards organisation.
  - any Input Parameters that need to be set for the Measurement Method, such as the address of the Measurement Peer
  - if the device with the MA has multiple interfaces, then the interface to use
  - optionally, a Cycle-ID
  - a name for this Measurement Task configuration
- configuration of all the Report Channels, each of which needs:
  - the address of the Collector, for instance its URL
  - the timing of when to report Measurement Results, for example every hour or immediately
  - security for sending the Report, for example the X.509 certificate
  - a name for this Report Channel
- the set of periodic Measurement Schedules, each of which needs:

    * the name of one or several Measurement Task configurations

    * the timing of when the Measurement Tasks are to be performed. Possible types of timing are periodic and calendar-based periodic

    * the name of a Report Channel or Channels on which to report the Measurement Results

    * a name for this Measurement Schedule

- the set of one-off Measurement Schedules, each of which needs the same items as for a periodic Measurement Schedule, except that the possible types of timing are one-off immediate and one-off at a future time.

A single Instruction message contains one, two, three or all four of the above elements. This allows the different elements to be updated independently at different times and intervals, for example it is likely that the periodic Measurement Schedule will be updated more often than the other elements.

Note that an Instruction message replaces (rather than adds to) those elements that it includes. For example, if the message includes (only) a periodic Measurement Schedule, then that replaces the old periodic Measurement Schedule but does not alter the configuration of the Measurement Tasks and Report Channels.

Periodic Measurement Schedules contain the name of one or several Measurement Task configurations that are to be carried out on a recurring basis, whilst one-off Measurement Schedules contain non-recurring Measurement Tasks. One-off and periodic Measurement Schedules are kept separate so that the Controller can instruct the MA to perform an ad hoc Measurement Task (for instance to help isolate a fault) without having to re-notify the MA about the periodic Measurement Schedule.

Note that the Instruction informs the MA; the Control Protocol does not allow the MA to negotiate, as this would add complexity to the MA, Controller and Control Protocol for little benefit.

The MA can inform the Controller about a Failure. There are two broad categories of failure: (1) the MA cannot action the Instruction (for example, it doesn't include a parameter that is mandatory for the requested Measurement Method; or it is missing details of the target Collector). (2) the MA cannot execute the Measurement Task or deliver the Report (for example, the MA unexpectedly has no spare CPU cycles; or the Collector is not responding). Note that it is not

considered a failure if a Measurement Task (correctly) doesn't start; for example if the MA detects cross-traffic, this is reported to the Collector in the normal manner. Note also that the MA does not inform the Controller about normal operation of its Measurement Tasks and Reports.

In the Figure, ACK means that the message has been delivered successfully.

Finally, note that the MA doesn't do a 'safety check' with the Controller (that it should still continue with the requested Measurement Tasks) - nor does it inform the Controller about Measurement Tasks starting and stopping. It simply carries out the Measurement Tasks as instructed, unless it gets an updated Instruction.

The LMAP WG will define a Control Protocol and its associated Data Model that implements the Protocol & Information Model. This may be a simple instruction-response protocol.

5.2.1. Measurement Suppression

Measurement Suppression is used if the measurement system wants to eliminate inessential traffic, because there is some unexpected network issue for example. The Controller instructs the MA to temporarily not begin new Active Measurement Tasks. By default, suppression applies to all Active Measurement Tasks, starts immediately and continues until an un-suppress message is received. Optionally the suppress message may include:

- a set of Active Measurement Tasks to suppress; the others are not suppressed. For example, a particular Measurement Task may be overloading a Measurement Peer.

- a set of Measurement Schedules to suppress; the others are not suppressed. For example, suppose the measurement system has defined two Schedules, one with the most critical Active Measurement Tasks and the other with less critical ones that create a lot of traffic, then it may only want to suppress the second.

- a start time, at which suppression begins

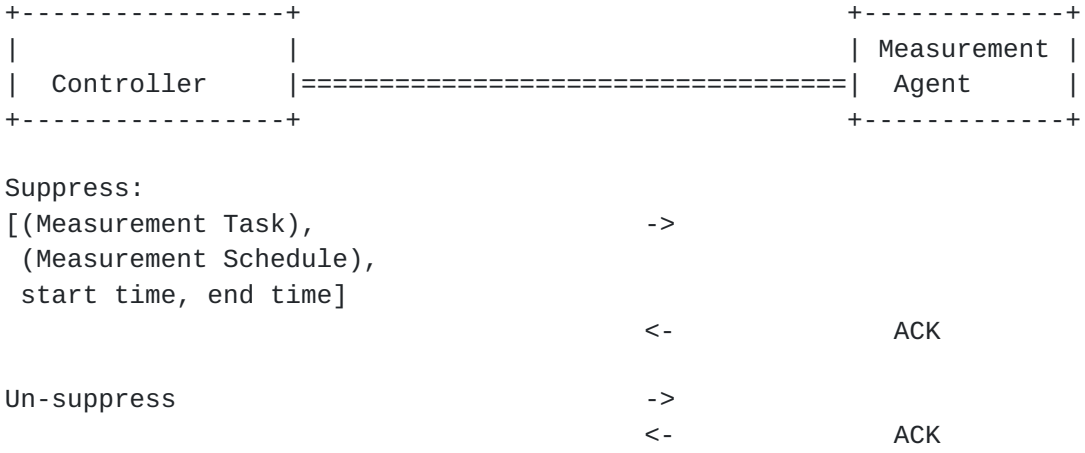
- an end time, at which suppression ends.

It is not standardised what the impact of Suppression is on:

- Passive Measurement Tasks; since they do not create any Active Measurement Traffic there is no need to suppress them, however it may be simpler for an implementation to do so

- on-going Active Measurement Tasks; see Section 5.3

Note that Suppression is not intended to permanently stop a Measurement Task (instead, the Controller should send a new Measurement Schedule), nor to permanently disable a MA (instead, some kind of management action is suggested).



5.3. Starting and stopping Measurement Tasks

The LMAP WG is neutral to what the actual Measurement Task is. The WG does not define a generic start and stop process, since the correct approach depend on the particular Measurement Task; the details are defined as part of each Measurement Method, and hence potentially by the IPPM WG. This section provides some general hints.

Once the MA gets its Measurement and Report Schedules from its Controller then it acts autonomously, in terms of operation of the Measurement Tasks and reporting of the result. One implication is that the MA initiates Measurement Tasks. As an example, for the common case where the MA is on a home gateway, the MA initiates a 'download speed test' by asking a Measurement Peer to send the file.

Many Active Measurement Tasks begin with a pre-check before the test traffic is sent. Action could include:

- the MA checking that there is no cross-traffic; in other words, a check that the user isn't already sending traffic;

- the MA checking with the Measurement Peer that it can handle a new Measurement Task (in case the Measurement Peer is already handling many Measurement Tasks with other MAs);

- the first part of the Measurement Task consisting of traffic that probes the path to make sure it isn't overloaded.

It is possible that similar checks continue during the Measurement Task, especially one that is long-running and/or creates a lot of Active Measurement Traffic, which may be abandoned whilst in-progress. A Measurement Task could also be abandoned in response to a "suppress" message (see Section 5.2.1). Action could include:

- For 'upload' tests, the MA not sending traffic

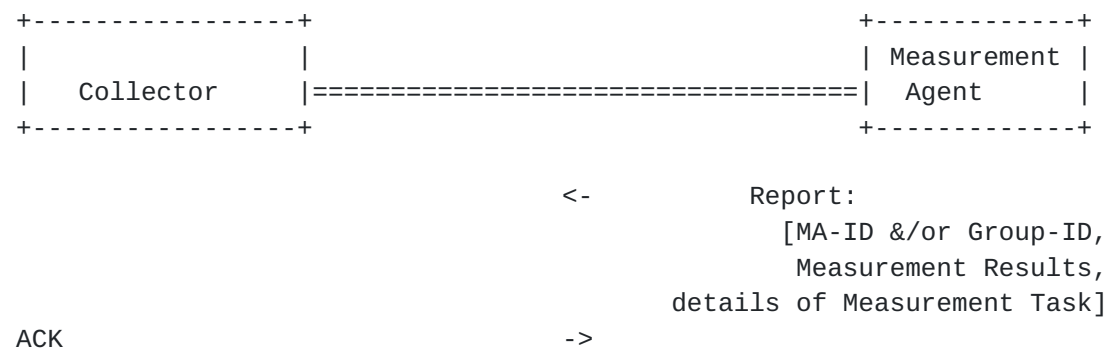
- For 'download' tests, the MA closing the TCP connection or sending a TWAMP Stop control message [RFC5357].

The Controller may want a MA to run the same Measurement Task indefinitely (for example, "run the 'upload speed' Measurement Task once an hour until further notice"). To avoid the MA generating traffic forever after a Controller has permanently failed, it is suggested that the Measurement Schedule includes a time limit ("run the 'upload speed' Measurement Task once an hour for the next 30 days") and that the Measurement Schedule is updated regularly (say, every 10 days).

{Comment: It is possible that the set of measurement schedules implies overlapping Measurement Tasks. It is not clear the best thing to do. Our current suggestion is to leave this to the protocol document.}

5.4. Report Protocol

The primary purpose of the Report Protocol is to allow a Measurement Agent to report its Measurement Results to a Collector, and the context in which they were obtained.


Collector
Measurement Agent
<- Report:
[MA-ID &/or Group-ID,
Measurement Results,
details of Measurement Task]
ACK ->
The Report contains:
o the MA-ID or a Group-ID (to anonymise results)
o the actual Measurement Results, including the time they were measured
o the details of the Measurement Task (to avoid the Collector having to ask the Controller for this information later)
The MA sends Reports as defined by the Report Channel in the Controller's Instruction. It is possible that the Instruction tells the MA to report the same Results to more than one Collector, or to report a different subset of Results to different Collectors. It is also possible that a Measurement Task may create two (or more) Measurement Results, which could be reported differently (for example, one Result could be reported periodically, whilst the second Result could be an alarm that is created as soon as the measured value of the Metric crosses a threshold and that is reported immediately).
Optionally, a Report is not sent when there are no Measurement Results.
In the initial LMAP Information Model and Report Protocol, for simplicity we assume that all Measurement Results are reported as-is, but allow extensibility so that a measurement system (or perhaps a second phase of LMAP) could allow a MA to pre-process Measurement Results before it reports them. Potential examples of pre-processing by the MA are:
o labelling, or perhaps not including, Measurement Results impacted by, for instance, cross-traffic or the Measurement Peer being busy
o not reporting the Measurement Results if the MA believes that they are invalid

- detailing when suppression started and ended
- filtering outlier Results
- calculating some statistic like average (beyond that defined by the Measurement Task itself)

The measurement system may define what happens if a Collector unexpectedly does not hear from a MA, for example the Controller could send a fresh Report Schedule to the MA.

The LMAP WG will define a Report Protocol and its associated Data Model that implements the Information Model and protocol model. This may be a simple instruction-response protocol.

5.5. Operation of LMAP over the underlying transport protocol

The above sections have described LMAP's protocol model. The LMAP working group will also specify how it operates over an existing protocol, to be selected, for example REST-style HTTP(S). It is also possible that a different choice is made for the Control and Report Protocols, for example NETCONF-YANG and IPFIX respectively. It is even possible that a different choice could be made for Suppression and for other Instruction messages.

For the Control Protocol, the underlying transport protocol could be:

- a 'push' protocol (that is, from the Controller to the MA)
- a multicast protocol (from the Controller to a group of MAs)
- a 'pull' protocol. The MA periodically checks with Controller if the Instruction has changed and pulls a new Instruction if necessary. A pull protocol seems attractive for a MA behind a NAT (as is typical for a MA on an end-user's device), so that it can initiate the communications. A pull mechanism is likely to require the MA to be configured with how frequently it should check in with the Controller, and perhaps what it should do if the Controller is unreachable after a certain number of attempts.
- a hybrid protocol. In addition to a pull protocol, the Controller can also push an alert to the MA that it should immediately pull a new Instruction.

For the Report Protocol, the underlying transport protocol could be:

- a 'push' protocol (that is, from the MA to the Collector)

- perhaps supplemented by the ability for the Collector to 'pull' Measurement Results from a MA.

5.6. Items beyond the scope of the LMAP Protocol Model

There are several potential interactions between LMAP elements that are out of scope of definition by the LMAP WG:

1. It does not define a coordination process between MAs. Whilst a measurement system may define coordinated Measurement Schedules across its various MAs, there is no direct coordination between MAs.

2. It does not define interactions between the Collector and Controller. It is quite likely that there will be such interactions, optionally intermediated by the data analysis tools. For example if there is an "interesting" Measurement Result then the measurement system may want to trigger extra Measurement Tasks that explore the potential cause in more detail.

3. It does not define coordination between different measurement systems. For example, it does not define the interaction of a MA in one measurement system with a Controller or Collector in a different measurement system. Whilst it is likely that the Control and Report Protocols could be re-used or adapted for this scenario, any form of coordination between different organisations involves difficult commercial and technical issues and so, given the novelty of large-scale measurement efforts, any form of inter-organisation coordination is outside the scope of the LMAP WG. Note that a single MA is instructed by a single Controller and is only in one measurement system.

   * An interesting scenario is where a home contains two independent MAs, for example one controlled by a regulator and one controlled by an ISP. Then the Active Measurement Traffic of one MA is treated by the other MA just like any other user traffic.

4. It does not consider how to prevent a malicious party "gaming the system". For example, where a regulator is running a measurement system in order to benchmark operators, a malicious operator could try to identify the broadband lines that the regulator was measuring and prioritise that traffic. It is assumed this is a policy issue and would be dealt with through a code of conduct for instance.

5. It does not define how to analyse Measurement Results, including how to interpret missing Results.

6. It does not specifically define a enduser-controlled measurement system, see sub-section 5.6.1.

5.6.1. Enduser-controlled measurement system

The WG concentrates on the cases where an ISP or a regulator runs the measurement system. However, we expect that LMAP functionality will also be used in the context of an enduser-controlled measurement system. There are at least two ways this could happen (they have various pros and cons):

1. an enduser could somehow request the ISP- (or regulator-) run measurement system to test his/her line. The ISP (or regulator) Controller would then send an Instruction to the MA in the usual LMAP way. Note that a user can't directly initiate a Measurement Task on an ISP- (or regulator-) controlled MA.

2. an enduser could deploy their own measurement system, with their own MA, Controller and Collector. For example, the user could implement all three functions onto the same enduser-owned end device, perhaps by downloading the functions from the ISP or regulator. Then the LMAP Control and Report Protocols do not need to be used, but using LMAP's Information Model would still be beneficial. The Measurement Peer could be in the home gateway or outside the home network; in the latter case the Measurement Peer is highly likely to be run by a different organisation, which raises extra privacy considerations.

In both cases there will be some way for the user to initiate the Measurement Task(s). The mechanism is out-of-scope of the LMAP WG, but could include the user clicking a button on a GUI or sending a text message. Presumably the user will also be able to see the Measurement Results, perhaps summarised on a webpage. It is suggested that these interfaces conform to the LMAP guidance on the privacy in Section 8.

6. Deployment considerations

6.1. Controller

The Controller should understand both the MA's LMAP Capabilities (for instance what Measurement Methods it can perform) and about the MA's other capabilities like processing power and memory. This allows the Controller to make sure that the Measurement Schedule of Measurement

Tasks and the Reporting Schedule are sensible for each MA that it Instructs.

An Instruction is likely to include several Measurement Tasks. Typically these run at different times, but it is also possible for them to run at the same time, if the Controller is sure that one Task will not affect the Results of another Task.

The Controller should ensure that the Active Measurement Tasks do not have an adverse effect on the end user. Typically Tasks, especially those that generate a substantial amount of traffic, will include a pre-check that the user isn't already sending traffic (Section 5.3). Another consideration is whether Active Measurement Traffic will impact a Subscriber's bill or traffic cap.

The different elements of the Instruction can be updated independently. For example, the Measurement Tasks could be configured with different Input Parameters whilst keeping the same Measurement Schedule. In general this should not create any issues, since Measurement Methods should be defined so their fundamental nature does not change for a new value of Input Parameter. There could be a problem if, for example, a Measurement Task involving a 1kB file upload could be changed into a 1GB file upload.

A measurement system may have multiple Controllers (but note the overriding principle that a single MA is instructed by a single Controller at any point in time (Section 4.2)). For example, there could be different Controllers for different types of MA (home gateways, tablets) or locations (Ipswich, Edinburgh), for load balancing or to cope with failure of one Controller. One possibility is that Bootstrapping involves an initial Controller, whose role is simply to inform the MA how to contact its actual Controller.

6.2. Measurement Agent

The Measurement Agent could take a number of forms: a dedicated probe, software on a PC, embedded into an appliance, or even embedded into a gateway. A single site (home, branch office etc.) that is participating in a measurement could make use of one or multiple Measurement Agents in a single measurement. If the site is multi homed there might be a Measurement Agent per interface.

The Measurement Agent could be deployed in a variety of locations. Not all deployment locations are available to every kind of Measurement Agent. There are also a variety of limitations and trade-offs depending on the final placement. The next sections outline some of the locations a Measurement Agent may be deployed. This is not an exhaustive list and combinations may also apply.

If the Instruction includes several Measurement Tasks, these could be scheduled to run at different times or possibly at the same time - some Tasks may be compatible, in that they do not affect each other's Results, whilst with others great care would need to be taken.

The measurement system also needs to consider carefully how to interpret missing Results; for example, if the missing Results are ignored and the lack of a Report is caused by its broadband being broken, then the estimate of overall performance, averaged across all MAs, would be too optimistic.

6.2.1. Measurement Agent embedded in site gateway

A Measurement Agent embedded with the site gateway, for example a home router or the edge router of a branch office in a managed service environment, is one of better places the Measurement Agent could be deployed. All site-to-ISP traffic would traverse through the gateway and passive measurements could easily be performed. Similarly, due to this user traffic visibility, an Active Measurements Task could be rescheduled so as not to compete with user traffic. Generally NAT and firewall services are built into the gateway, allowing the Measurement Agent the option to offer its Controller facing management interface outside of the NAT/firewall. This placement of the management interface allows the Controller to unilaterally contact the Measurement Agent for instructions. However, if the site gateway is owned and operated by the service provider, the Measurement Agent will generally not be directly available for over the top providers, the regulator, end users or enterprises.

6.2.2. Measurement Agent embedded behind site NAT /Firewall

The Measurement Agent could also be embedded behind a NAT, a firewall, or both. In this case the Controller may not be able to unilaterally contact the Measurement Agent unless either static port forwarding configuration or firewall pin holing is configured, and might not always be possible. It would require user intervention or pre-provisioning by the operator via a mechanisms such as TR-069. The Measurement Agent may originate a session towards the Controller and maintain the session for bidirectional communications. This would alleviate the need to have user intervention on the gateway, but would reduce the overall saleability of the Controller as it would have to maintain a higher number of active sessions. That said, sending keepalives to prop open the firewall could serve a dual purpose in testing network reachability for the Measurement Agent. An alternative would be to use a protocol such as UPnP or PCP [RFC6887] to control the NAT/firewall if the gateway supports this kind of control.

6.2.3. Measurement Agent in a multi-homed site

A broadband site may be multi-homed. For example, the site may be connected to multiple broadband ISPs, perhaps for redundancy or load-sharing, or have both wired and wireless broadband connectivity. It may also be helpful to think of dual stack IPv4 and IPv6 broadband devices as multi-homed. In these cases, there needs to be clarity on which network connectivity option is being measured. Sometimes this is easily resolved by the location of the MA itself. For example, if the MA is built into the gateway (and the gateway only has a single WAN side interface), there is little confusion or choice. However, for multi-homed gateways or devices behind the gateway(s) of multi-homed sites it would be preferable to explicitly select the network to measure ([RFC5533]) but the network measured should be included in the Measurement Result. Section 3.2 of [I-D.ietf-homenet-arch] describes dual-stack and multi-homing topologies that might be encountered in a home network (which is generally a broadband connected site). The Multiple Interfaces (mif) working group covers cases where hosts are either directly attached to multiple networks (physical or virtual) or indirectly (multiple default routers, etc.). [RFC6419] provides the current practices of multi-interfaces hosts today. As one aim is for a MA is to measure the end user's quality of experience, it is important to understand the current practices.

6.3. Measurement Peer

A Measurement Peer participates in Active Measurement Tasks. It may have specific functionality to enable it to participate in a particular Measurement Method. On the other hand, other Measurement Methods may require no special functionality, for example if the Measurement Agent sends a ping to example.com then the server at example.com plays the role of a Measurement Peer.

A device may participate in some Measurement Tasks as a Measurement Agent and in others as a Measurement Peer.

7. Security considerations

The security of the LMAP framework should protect the interests of the measurement operator(s), the network user(s) and other actors who could be impacted by a compromised measurement deployment. The measurement system must secure the various components of the system from unauthorised access or corruption.

We assume that each Measurement Agent (MA) will receive its Instructions from a single organisation, which operates the Controller. These Instructions must be authenticated (to ensure that they come from the trusted Controller), checked for integrity (to

ensure no-one has tampered with them) and not vulnerable to replay attacks. If a malicious party can gain control of the MA they can use it to launch DoS attacks at targets, reduce the end user's quality of experience and corrupt the Measurement Results that are reported to the Collector. By altering the Measurement Tasks and/or the address that Results are reported to, they can also compromise the confidentiality of the network user and the MA environment (such as information about the location of devices or their traffic).

Reporting by the MA must also be secured to maintain confidentiality. The results must be encrypted such that only the authorised Collector can decrypt the results to prevent the leakage of confidential or private information. In addition it must be authenticated that the results have come from the expected MA and that they have not been tampered with. It must not be possible to fool a MA into injecting falsified data into the measurement platform or to corrupt the results of a real MA. The results must also be held and processed securely after collection and analysis.

Availability should also be considered. While the loss of some MAs may not be considered critical, the unavailability of the Collector could mean that valuable business data or data critical to a regulatory process is lost. Similarly, the unavailability of a Controller could mean that the MAs do not operate a correct Measurement Schedule.

A malicious party could "game the system". For example, where a regulator is running a measurement system in order to benchmark operators, an operator could try to identify the broadband lines that the regulator was measuring and prioritise that traffic. This potential issue is currently handled by a code of conduct. It is outside the scope of the LMAP WG to consider the issue.

8. Privacy Considerations for LMAP

The LMAP Working Group will consider privacy as a core requirement and will ensure that by default the Control and Report Protocols operate in a privacy-sensitive manner and that privacy features are well-defined.

This section provides a set of privacy considerations for LMAP. This section benefits greatly from the timely publication of [RFC6973]. Privacy and security (Section 7) are related. In some jurisdictions privacy is called data protection.

We begin with a set of assumptions related to protecting the sensitive information of individuals and organisations participating in LMAP-orchestrated measurement and data collection.

8.1. Categories of Entities with Information of Interest

LMAP protocols need to protect the sensitive information of the following entities, including individuals and organisations who participate in measurement and collection of results.

- Individual Internet users: Persons who utilise Internet access services for communications tasks, according to the terms of service of a service agreement. Such persons may be a service Subscriber, or have been given permission by the Subscriber to use the service.

- Internet service providers: Organisations who offer Internet access service subscriptions, and thus have access to sensitive information of individuals who choose to use the service. These organisations desire to protect their Subscribers and their own sensitive information which may be stored in the process of performing Measurement Tasks and collecting and Results.

- Regulators: Public authorities responsible for exercising supervision of the electronic communications sector, and which may have access to sensitive information of individuals who participate in a measurement campaign. Similarly, regulators desire to protect the participants and their own sensitive information.

- Other LMAP system operators: Organisations who operate measurement systems or participate in measurements in some way.

Although privacy is a protection extended to individuals, we include discussion of ISPs and other LMAP system operators in this section. These organisations have sensitive information involved in the LMAP system, and many of the same dangers and mitigations are applicable. Further, the ISPs store information on their Subscribers beyond that used in the LMAP system (for instance billing information), and there should be a benefit in considering all the needs and potential solutions coherently.

8.2. Examples of Sensitive Information

This section gives examples of sensitive information which may be measured or stored in a measurement system, and which is to be kept private by default in the LMAP core protocols.

Examples of Subscriber or authorised Internet user sensitive information:

- Sub-IP layer addresses and names (MAC address, base station ID, SSID)
- IP address in use
- Personal Identification (real name)
- Location (street address, city)
- Subscribed service parameters
- Contents of traffic (activity, DNS queries, destinations, equipment types, account info for other services, etc.)
- Status as a study volunteer and Schedule of (Active) Measurement Tasks

Examples of Internet Service Provider sensitive information:

- Measurement device identification (equipment ID and IP address)
- Measurement Instructions (choice of measurements)
- Measurement Results (some may be shared, others may be private)
- Measurement Schedule (exact times)
- Network topology (locations, connectivity, redundancy)
- Subscriber billing information, and any of the above Subscriber information known to the provider.
- Authentication credentials (such as certificates)

Other organisations will have some combination of the lists above. The LMAP system would not typically expose all of the information above, but could expose a combination of items which could be correlated with other pieces collected by an attacker (as discussed in the section on Threats below).

8.3. Key Distinction Between Active and Passive Measurement Tasks

Passive and Active Measurement Tasks raise different privacy issues.

Passive Measurement Tasks are conducted on a user's traffic, such that sensitive information is present and stored in the measurement system (however briefly this storage may be). We note that some authorities make a distinction on time of storage, and information

that is kept only temporarily to perform a communications function is not subject to regulation (for example, active queue management, deep packet inspection). Passive Measurement Tasks could reveal all the websites a Subscriber visits and the applications and/or services they use.

Active Measurement Tasks are conducted on traffic which is created specifically for the purpose. Even if a user host generates Active Measurement Traffic, there is significantly limited sensitive information about the Subscriber present and stored in the measurement system compared to the passive case, as follows:

- IP address in use (and possibly sub-IP addresses and names)

- Status as a study volunteer and Schedule of Active Measurement Tasks

On the other hand, for a service provider the sensitive information like Measurement Results is the same for Passive and Active Measurement Tasks.

From the Subscriber perspective, both Active and Passive Measurement Tasks potentially expose the description of Internet access service and specific service parameters, such as subscribed rate and type of access.

8.4. Privacy analysis of the Communications Models

This section examines each of the protocol exchanges described at a high level in Section 5 and some example Measurement Tasks, and identifies specific sensitive information which must be secured during communication for each case. With the protocol-related sensitive information identified, we have can better consider the threats described in the following section.

From the privacy perspective, all entities participating in LMAP protocols can be considered "observers" according to the definition in [RFC6973]. Their stored information potentially poses a threat to privacy, especially if one or more of these functional entities has been compromised. Likewise, all devices on the paths used for control, reporting, and measurement are also observers.

8.4.1. MA Bootstrapping

Section 5.1 provides the communication model for the Bootstrapping process.

Although the specification of mechanisms for Bootstrapping the MA are beyond the LMAP scope, designers should recognize that the Bootstrapping process is extremely powerful and could cause an MA to join a new or different LMAP system with a different Controller and Collector, or simply install new Measurement Methods (for example to passively record DNS queries). A Bootstrap attack could result in a breach of the LMAP system with significant sensitive information exposure depending on the capabilities of the MA, so sufficient security protections are warranted.

The Bootstrapping process provides sensitive information about the LMAP system and the organisation that operates it, such as

- Initial Controller IP address or FQDN
- Assigned Controller IP address or FQDN
- Security certificates and credentials

During the Bootstrap process, the MA receives its MA-ID which is a persistent pseudonym for the Subscriber in the case that the MA is located at a service demarcation point. Thus, the MA-ID is considered sensitive information, because it could provide the link between Subscriber identification and Measurements Results.

Also, the Bootstrap process could assign a Group-ID to the MA. The specific definition of information represented in a Group-ID is to be determined, but several examples are envisaged including use as a pseudonym for a set of Subscribers, a class of service, an access technology, or other important categories. Assignment of a Group-ID enables anonymisation sets to be formed on the basis of service type/ grade/rates. Thus, the mapping between Group-ID and MA-ID is considered sensitive information.

8.4.2. Controller <-> Measurement Agent

The high-level communication model for interactions between the LMAP Controller and Measurement Agent is illustrated in Section 5.2. The primary purpose of this exchange is to authenticate and task a Measurement Agent with Measurement Instructions, which the Measurement Agent then acts on autonomously.

Primarily IP addresses and pseudonyms (MA-ID, Group-ID) are exchanged with a capability request, then measurement-related information of interest such as the parameters, schedule, metrics, and IP addresses of measurement devices. Thus, the measurement Instruction contains sensitive information which must be secured. For example, the fact that an ISP is running additional measurements beyond the set

reported externally is sensitive information, as are the additional Measurements Tasks themselves. The Measurement Schedule is also sensitive, because an attacker intending to bias the results without being detected can use this information to great advantage.

An organisation operating the Controller having no service relationship with a user who hosts the Measurement Agent *could* gain real-name mapping to a public IP address through user participation in an LMAP system (this applies to the Measurement Collection protocol, as well).

8.4.3. Collector <-> Measurement Agent

The high-level communication model for interactions between the Measurement Agent and Collector is illustrated in Section 5.4. The primary purpose of this exchange is to authenticate and collect Measurement Results from a MA, which the MA has measured autonomously and stored.

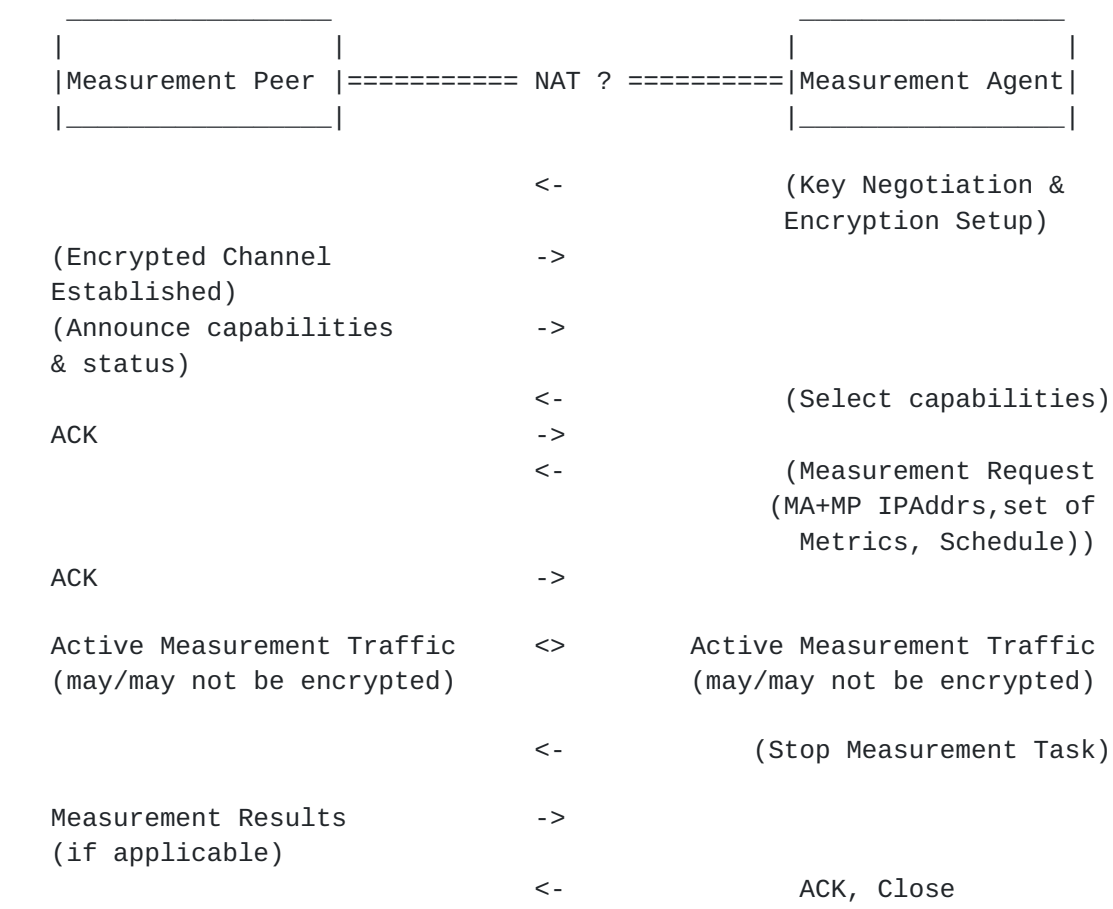
The Measurement Results are the additional sensitive information included in the Collector-MA exchange. Organisations collecting LMAP measurements have the responsibility for data control. Thus, the Results and other information communicated in the Collector protocol must be secured.

8.4.4. Measurement Peer <-> Measurement Agent

Although the specification of the mechanisms for an Active Measurement Task is beyond the scope of LMAP, it raises potential privacy issues. The high-level communications model below illustrates the various exchanges to execute Active Measurement Tasks and store the Results.

We note the potential for additional observers in the figures below by indicating the possible presence of a NAT, which has additional significance to the protocols and direction of initiation.

The various messages are optional, depending on the nature of the Active Measurement Task. It may involve sending Active Measurement Traffic from the Measurement Peer to MA, MA to Measurement Peer, or both.



This exchange primarily exposes the IP addresses of measurement devices and the inference of measurement participation from such traffic. There may be sensitive information on key points in a service provider's network included. There may also be access to measurement-related information of interest such as the Metrics, Schedule, and intermediate results carried in the Active Measurement Traffic (usually a set of timestamps).

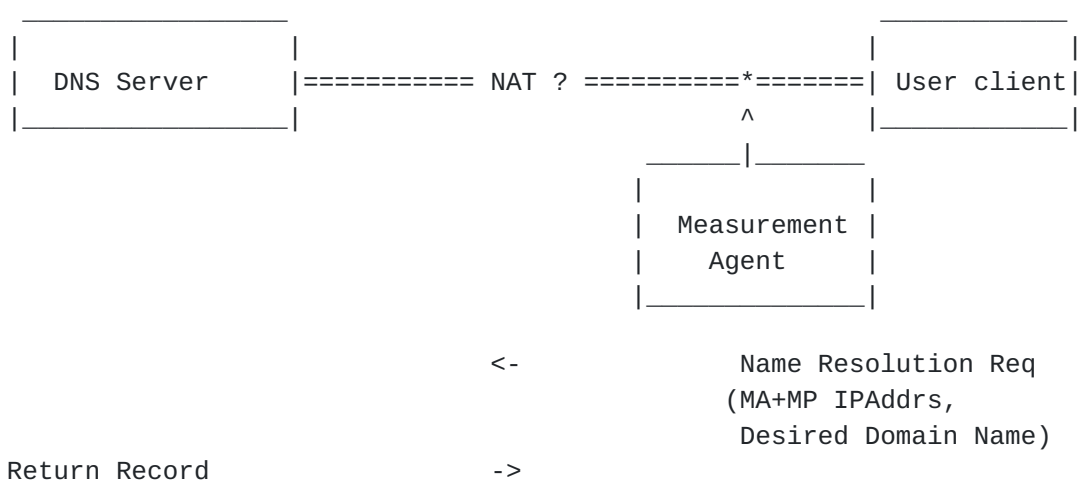
If the Active Measurement Traffic is unencrypted, as found in many systems today, then both timing and limited results are open to on-path observers.

8.4.5. Passive Measurement Agent

Although the specification of the mechanisms for a Passive Measurement Task is beyond the scope of LMAP, it raises potential privacy issues.

The high-level communications model below illustrates the collection of user information of interest with the Measurement Agent performing the monitoring and storage of the Results. This particular exchange

is for passive measurement of DNS Response Time, which most frequently uses UDP transport.



This exchange primarily exposes the IP addresses of measurement devices and the intent to communicate with or access the services of "Domain Name". There may be information on key points in a service provider's network, such as the address of one of its DNS servers. The Measurement Agent may be embedded in the user host, or it may be located in another device capable of observing user traffic.

In principle, any of the user sensitive information of interest (listed above) can be collected and stored in the passive monitoring scenario and so must be secured.

It would also be possible for a Measurement Agent to source the DNS query itself. But then, as with any active measurement task, there are few privacy concerns.

8.4.6. Storage and Reporting of Measurement Results

Although the mechanisms for communicating results (beyond the initial Collector) are beyond the LMAP scope, there are potential privacy issues related to a single organisation's storage and reporting of Measurement Results. Both storage and reporting functions can help to preserve privacy by implementing the mitigations described below.

8.5. Threats

This section indicates how each of the threats described in [RFC6973] apply to the LMAP entities and their communication and storage of "information of interest".

8.5.1. Surveillance

Section 5.1.1 of [RFC6973] describes Surveillance as the "observation or monitoring of and individual's communications or activities." Hence all Passive Measurement Tasks are a form of surveillance, with inherent risks.

Active Measurement Methods which avoid periods of user transmission indirectly produce a record of times when a subscriber or authorised user has used their network access service.

Active Measurement Methods may also utilise and store a Subscriber's currently assigned IP address when conducting measurements that are relevant to a specific Subscriber. Since the Measurement Results are time-stamped, they could provide a record of IP address assignments over time.

Either of the above pieces of information could be useful in correlation and identification, described below.

8.5.2. Stored Data Compromise

Section 5.1.2 of [RFC6973] describes Stored Data Compromise as resulting from inadequate measures to secure stored data from unauthorised or inappropriate access. For LMAP systems this includes deleting or modifying collected measurement records, as well as data theft.

The primary LMAP entity subject to compromise is the repository, which stores the Measurement Results; extensive security and privacy threat mitigations are warranted. The Collector and MA also store sensitive information temporarily, and need protection. The communications between the local storage of the Collector and the repository is beyond the scope of the LMAP work at this time, though this communications channel will certainly need protection as well as the mass storage itself.

The LMAP Controller may have direct access to storage of Subscriber information (location, billing, service parameters, etc.) and other information which the controlling organisation considers private, and again needs protection.

8.5.3. Correlation and Identification

Sections 5.2.1 and 5.2.2 of [RFC6973] describes Correlation as combining various pieces of information to obtain desired characteristics of an individual, and Identification as using this process to infer identity.

The main risk is that the LMAP system could unwittingly provide a key piece of the correlation chain, starting with an unknown Subscriber's IP address and another piece of information. For example, a Subscriber utilised Internet access from 2000 to 2310 UTC, because the Active Measurement Tasks were deferred, or sent a name resolution for www.example.com at 2300 UTC.

8.5.4. Secondary Use and Disclosure

Sections 5.2.3 and 5.2.4 of [RFC6973] describes Secondary Use as unauthorised utilisation of an individual's information for a purpose the individual did not intend, and Disclosure is when such information is revealed causing other's notions of the individual to change, or confidentiality to be violated.

Passive Measurement Tasks are a form of Secondary Use, and the Subscribers' permission and the measured ISP's permission should be obtained beforehand. Although user traffic is only indirectly involved, the Measurement Results from Active Measurement Tasks provide some limited information about the Subscriber/ISP and could be used for Secondary Uses. For example, the use of the Results in unauthorised marketing campaigns would qualify as Secondary Use.

8.6. Mitigations

This section examines the mitigations listed in section 6 of [RFC6973] and their applicability to LMAP systems. Note that each section in [RFC6973] identifies the threat categories that each technique mitigates.

8.6.1. Data Minimisation

Section 6.1 of [RFC6973] encourages collecting and storing the minimal information needed to perform a task.

There are two levels of information needed for LMAP results to be useful for a specific task: troubleshooting and general results reporting.

For general results, the results can be aggregated into large categories (the month of March, all subscribers West of the Mississippi River). In this case, all individual identifications (including IP address of the MA) can be excluded, and only relevant results are provided. However, this implies a filtering process to reduce the information fields, because greater detail was needed to conduct the Measurement Tasks in the first place.

For troubleshooting, so that a network operator or end user can identify a performance issue or failure, potentially all the network information (IP addresses, equipment IDs, location), Measurement Schedule, service configuration, Measurement Results, and other information may assist in the process. This includes the information needed to conduct the Measurements Tasks, and represents a need where the maximum relevant information is desirable, therefore the greatest protections should be applied.

We note that a user may give temporary permission for Passive Measurement Tasks to enable detailed troubleshooting, but withhold permission for them in general. Here the greatest breadth of sensitive information is potentially exposed, and the maximum privacy protection must be provided.

For MAs with access to the sensitive information of users (e.g., within a home or a personal host/handset), it is desirable for the results collection to minimise the data reported, but also to balance this desire with the needs of troubleshooting when a service subscription exists between the user and organisation operating the measurements.

For passive measurements where the MA reports flow information to the Collector, the Collector may perform pre-storage minimisation and other mitigations (below) to help preserve privacy.

8.6.2. Anonymity

Section 6.1.1 of [RFC6973] describes a way in which anonymity is achieved: "there must exist a set of individuals that appear to have the same attributes as the individual", defined as an "anonymity set".

Experimental methods for anonymisation of user identifiable data applicable to Passive Measurement Methods have been identified in [RFC6235]. However, the findings of several of the same authors is that "there is increasing evidence that anonymisation applied to network trace or flow data on its own is insufficient for many data protection applications as in [Bur10]."

Essentially, the details of passive measurement tasks can only be accessed by closed organisations, and unknown injection attacks are always less expensive than the protections from them. However, some forms of summary may protect the user's sensitive information sufficiently well, and so each Metric must be evaluated in the light of privacy.

The methods in [RFC6235] could be applied more successfully in Active Measurement Methods, where there are protections from injection attack. The successful attack would require breaking the integrity protection of the LMAP Reporting Protocol and injecting Measurement Results (known fingerprint, see section 3.2 of [RFC6973]) for inclusion with the shared and anonymised results, then fingerprinting those records to ascertain the anonymisation process.

Beside anonymisation of measured Results for a specific user or provider, the value of sensitive information can be further diluted by summarising the results over many individuals or areas served by the provider. There is an opportunity enabled by forming anonymity sets [RFC6973] based on the reference path measurement points in [I-D.ietf-ippm-lmap-path]. For example, all measurements from the Subscriber device can be identified as "mp000", instead of using the IP address or other device information. The same anonymisation applies to the Internet Service Provider, where their Internet gateway would be referred to as "mp190".

8.6.3. Pseudonymity

Section 6.1.2 of [RFC6973] indicates that pseudonyms, or nicknames, are a possible mitigation to revealing one's true identity, since there is no requirement to use real names in almost all protocols.

A pseudonym for a measurement device's IP address could be an LMAP-unique equipment ID. However, this would likely be a permanent handle for the device, and long-term use weakens a pseudonym's power to obscure identity.

8.6.4. Other Mitigations

Data can be de-personalised by blurring it, for example by adding synthetic data, data-swapping, or perturbing the values in ways that can be reversed or corrected.

Sections 6.2 and 6.3 of [RFC6973] describe User Participation and Security, respectively.

Where LMAP measurements involve devices on the Subscriber's premises or Subscriber-owned equipment, it is essential to secure the Subscriber's permission with regard to the specific information that will be collected. The informed consent of the Subscriber (and, if different, the end user) is needed, including the specific purpose of the measurements. The approval process could involve showing the Subscriber their measured information and results before instituting periodic collection, or before all instances of collection, with the option to cancel collection temporarily or permanently.

It should also be clear who is legally responsible for data protection (privacy); in some jurisdictions this role is called the 'data controller'. It is good practice to time limit the storage of personal information.

Although the details of verification would be impenetrable to most subscribers, the MA could be architected as an "app" with open source-code, pre-download and embedded terms of use and agreement on measurements, and protection from code modifications usually provided by the app-stores. Further, the app itself could provide data reduction and temporary storage mitigations as appropriate and certified through code review.

LMAP protocols, devices, and the information they store clearly need to be secure from unauthorised access. This is the hand-off between privacy and security considerations (Section 7). The Data Controller has the (legal) responsibility to maintain data protections described in the Subscriber's agreement and agreements with other organisations.

9. IANA Considerations

There are no IANA considerations in this memo.

10. Acknowledgments

This document is a merger of three individual drafts: draft-eardley-lmap-terminology-02, draft-akhter-lmap-framework-00, and draft-eardley-lmap-framework-02.

Thanks to Juergen Schoenwaelder for his detailed review of the terminology. Thanks to Charles Cook for a very detailed review of -02.

Thanks to numerous people for much discussion, directly and on the LMAP list (apologies to those unintentionally omitted): Alan Clark, Alissa Cooper, Andrea Soppera, Barbara Stark, Benoit Claise, Brian Trammell, Charles Cook, Dave Thorne, Frode Soerensen, Greg Mirsky, Guangqing Deng, Jason Weil, Jean-Francois Tremblay, Jerome Benoit, Joachim Fabini, Juergen Schoenwaelder, Jukka Manner, Ken Ko, Michael Bugenhagen, Rolf Winter, Sam Crawford, Sharam Hakimi, Steve Miller, Ted Lemon, Timothy Carey, Vaibhav Bajpai, William Lupton.

Philip Eardley, Trevor Burbridge and Marcelo Bagnulo work in part on the Leone research project, which receives funding from the European Union Seventh Framework Programme [FP7/2007-2013] under grant agreement number 317647.

11. History

First WG version, copy of draft-folks-lmap-framework-00.

11.1. From -00 to -01

- new sub-section of possible use of Group-IDs for privacy
- tweak to definition of Control protocol
- fix typo in figure in S5.4

11.2. From -01 to -02

- change to INFORMATIONAL track (previous version had typo'd Standards track)
- new definitions for Capabilities Information and Failure Information
- clarify that diagrams show LMAP-level information flows. Underlying protocol could do other interactions, eg to get through NAT or for Collector to pull a Report
- add hint that after a re-boot should pause random time before re-register (to avoid mass calling event)
- delete the open issue "what happens if a Controller fails" (normal methods can handle)
- add some extra words about multiple Tasks in one Schedule
- clarify that new Schedule replaces (rather than adds to) and old one. Similarly for new configuration of Measurement Tasks or Report Channels.
- clarify suppression is temporary stop; send a new Schedule to permanently stop Tasks
- alter suppression so it is ACKed
- add un-suppress message
- expand the text on error reporting, to mention Reporting failures (as well as failures to action or execute Measurement Task & Schedule)
- add some text about how to have Tasks running indefinitely

- add that optionally a Report is not sent when there are no Measurement Results
- add that a Measurement Task may create more than one Measurement Result
- clarify /amend /expand that Reports include the "raw" Measurement Results - any pre-processing is left for lmap2.0
- add some cautionary words about what if the Collector unexpectedly doesn't hear from a MA
- add some extra words about the potential impact of Measurement Tasks
- clarified various aspects of the privacy section
- updated references
- minor tweaks

11.3. From -02 to -03

- alignment with the Information Model [I-D.burbridge-lmap-information-model] as this is agreed as a WG document
- One-off and periodic Measurement Schedules are kept separate, so that they can be updated independently
- Measurement Suppression in a separate sub-section. Can now optionally include particular Measurement Tasks &/or Schedules to suppress, and start/stop time
- for clarity, concept of Channel split into Control, Report and MA-to-Controller Channels
- numerous editorial changes, mainly arising from a very detailed review by Charles Cook
-

12. Informative References

[Bur10] Burkhart, M., Schatzmann, D., Trammell, B., and E. Boschi, "The Role of Network Trace anonymisation Under Attack", January 2010.

[Q1741] Q.1741.7, , "IMT-2000 references to Release 9 of GSM-evolved UMTS core network", http://www.itu.int/rec/T-REC-Q.1741.7/en, November 2011.

[RFC1035] Mockapetris, P., "Domain names - implementation and specification", STD 13, RFC 1035, November 1987.

[RFC4101] Rescorla, E. and IAB, "Writing Protocol Models", RFC 4101, June 2005.

[RFC4122] Leach, P., Mealling, M., and R. Salz, "A Universally Unique IDentifier (UUID) URN Namespace", RFC 4122, July 2005.

[RFC5357] Hedayat, K., Krzanowski, R., Morton, A., Yum, K., and J. Babiarz, "A Two-Way Active Measurement Protocol (TWAMP)", RFC 5357, October 2008.

[I-D.ietf-lmap-use-cases] Linsner, M., Eardley, P., Burbridge, T., and F. Sorensen, "Large-Scale Broadband Measurement Use Cases", draft-ietf-lmap-use-cases-01 (work in progress), December 2013.

[I-D.bagnulo-ippm-new-registry-independent] Bagnulo, M., Burbridge, T., Crawford, S., Eardley, P., and A. Morton, "A registry for commonly used metrics. Independent registries", draft-bagnulo-ippm-new-registry-independent-01 (work in progress), July 2013.

[I-D.ietf-homenet-arch] Chown, T., Arkko, J., Brandt, A., Troan, O., and J. Weil, "IPv6 Home Networking Architecture Principles", draft-ietf-homenet-arch-11 (work in progress), October 2013.

[RFC6419] Wasserman, M. and P. Seite, "Current Practices for Multiple-Interface Hosts", RFC 6419, November 2011.

[RFC6887] Wing, D., Cheshire, S., Boucadair, M., Penno, R., and P. Selkirk, "Port Control Protocol (PCP)", RFC 6887, April 2013.

[RFC5533] Nordmark, E. and M. Bagnulo, "Shim6: Level 3 Multihoming Shim Protocol for IPv6", RFC 5533, June 2009.

[I-D.burbridge-lmap-information-model]
Burbridge, T., Eardley, P., Bagnulo, M., and J. Schoenwaelder, "Information Model for Large-Scale Measurement Platforms (LMAP)", draft-burbridge-lmap-information-model-01 (work in progress), October 2013.

[RFC6235] Boschi, E. and B. Trammell, "IP Flow Anonymization Support", RFC 6235, May 2011.

[RFC6973] Cooper, A., Tschofenig, H., Aboba, B., Peterson, J., Morris, J., Hansen, M., and R. Smith, "Privacy Considerations for Internet Protocols", RFC 6973, July 2013.

[I-D.ietf-ippm-lmap-path]
Bagnulo, M., Burbridge, T., Crawford, S., Eardley, P., and A. Morton, "A Reference Path and Measurement Points for LMAP", draft-ietf-ippm-lmap-path-01 (work in progress), September 2013.

Authors' Addresses

Philip Eardley
British Telecom
Adastral Park, Martlesham Heath
Ipswich
ENGLAND

Email: philip.eardley@bt.com

Al Morton
AT&T Labs
200 Laurel Avenue South
Middletown, NJ
USA

Email: acmorton@att.com

Marcelo Bagnulo
Universidad Carlos III de Madrid
Av. Universidad 30
Leganes, Madrid 28911
SPAIN

Phone: 34 91 6249500
Email: marcelo@it.uc3m.es
URI: http://www.it.uc3m.es

Trevor Burbridge
British Telecom
Adastral Park, Martlesham Heath
Ipswich
ENGLAND

Email: trevor.burbridge@bt.com

Paul Aitken
Cisco Systems, Inc.
96 Commercial Street
Edinburgh, Scotland EH6 6LX
UK

Email: paitken@cisco.com

Aamer Akhter
Cisco Systems, Inc.
7025 Kit Creek Road
RTP, NC 27709
USA

Email: aakhter@cisco.com