GENI

Global Environment for Network Innovations

GENI Instrumentation and Measurement Architecture

Document ID: GENI-SE-IM-ARCH-0.1

March 8, 2010

Prepared by:
The GENI Project Office
BBN Technologies
10 Moulton Street
Cambridge, MA 02138 USA

Issued under NSF Cooperative Agreement CNS-0737890

TABLE OF CONTENTS

1 Document Scope

This section describes this document's purpose, its context within the overall GENI document tree, the set of related documents, and this document's revision history.

1.1 Purpose of this Document

This document provides an overview of the GENI suite structure.

It is a DRAFT, to be used for discussion..

1.2 Context for this Document

Figure 1-1. below shows the context for this document within GENI's overall document tree.

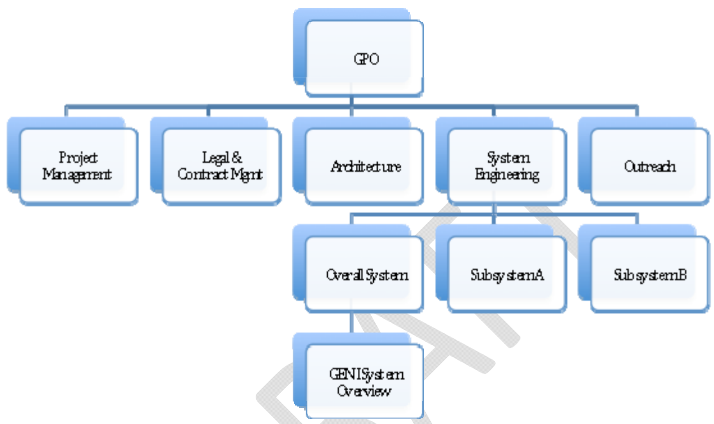

Figure 1-1. This Document within the GENI Document Tree.

1.3 Related Documents

The following documents of exact date listed are related to this document, and provide background information, requirements, etc., that are important for this document.

1.3.1 National Science Foundation (NSF) Documents

Document ID	Document Title and Issue Date
N / A	

1.3.2 GENI Documents

Document ID	Document Title and Issue Date
GENI-SE-SY-RQ-01.9	GENI System Requirements (DRAFT) January 16, 2009 http://groups.geni.net/geni/wiki/SysReqDoc
GENI-SE-SY-SO-02.0	GENI System Overview September 29, 2008 http://www.geni.net/docs/GENISysOvrvw092908.pdf
GENI-INF-PRO-S1-OV-1.12	GENI Spiral 1 Overview September 29, 2009 http://www.geni.net/docs/GENIS1Ovrvw092908.pdf
GENI-SE-CF-RQ-01.3	GENI Control Framework Requirements (DRAFT) January 13, 2009 http://groups.geni.net/geni/wiki/GeniControlFrameworkRequirements
GENI-SE-CF-PLGO-01.2	PlanetLab GENI Control Framework Overview (DRAFT) January 14, 2009 http://groups.geni.net/geni/wiki/PlanetLabGeniControlFrameworkOverview
GENI-SE-CF-PRGO-01.4	ProtoGENI Control Framework Overview (DRAFT) March 25, 2009 http://groups.geni.net/geni/wiki/ProtoGeniControlFrameworkOverview
GENI-SE-CF-ORGO-01.2	ORCA GENI Control Framework Overview (DRAFT) January 14, 2009 http://groups.geni.net/geni/wiki/OrcaGeniControlFrameworkOverview
GENI-SE-SY-TS-UC-LC-01.1	Lifecycle of a GENI Experiment (DRAFT) March 4, 2009 http://groups.geni.net/geni/wiki/ExperimentLifecycleDocument

1.3.3 Standards Documents

Document ID	Document Title and Issue Date
N / A	

1.3.4 Other Documents

Document ID	Document Title and Issue Date

1.4 Document Revision History

The following table provides the revision history for this document, summarizing the date at which it was revised, who revised it, and a brief summary of the changes. This list is maintained in reverse chronological order so the newest revision comes first in the list.

Revision	Date	Revised By	Summary of Changes
-0.1	3/8/10	H. Mussman	Early draft, for review within the GENI Instrumentation and Measurements Working Group

2 Introduction

The Global Environment for Network Innovations (GENI) is a suite of experimental network research infrastructure now being planned and prototyped.

To conduct useful and repeatable experiments, GENI researchers require extensive and reliable instrumentation and measurement capabilities to gather, analyze, present and archive Measurement Data (MD).

In addition, the GENI operations staff require extensive and reliable instrumentation and measurement capabilities to monitor and troubleshoot the GENI suite and its constituent entities.

This document presents the GENI Instrumentation and Measurement (I&M) architecture, comprising several key Instrumentation and Measurement (I&M) services:

- Measurement Orchestration (MO) service
- Measurement Point (MP) service
- Measurement Collection (MC) service
- Measurement Analysis and Presentation (MAP) service
- Measurement Data Archive (MDA) service

3 Configuration of Instrumentation and Measurement (I&M) Services

3.1 Basic Experiment Configuration

Researcher, who uses an Experiment Control (EC) service (tools) to gather resources into a particular "slice" (container of resources), configure and/or program them, and then manage them to setup and run an experiment.

Control Framework (CF), which allows a Researcher to discover, authorize and assign resources to a slice for an experiment.

Resources in Aggregates, which have been assigned to a researcher/slice for an experiment.

3.2 Basic I&M Configuration

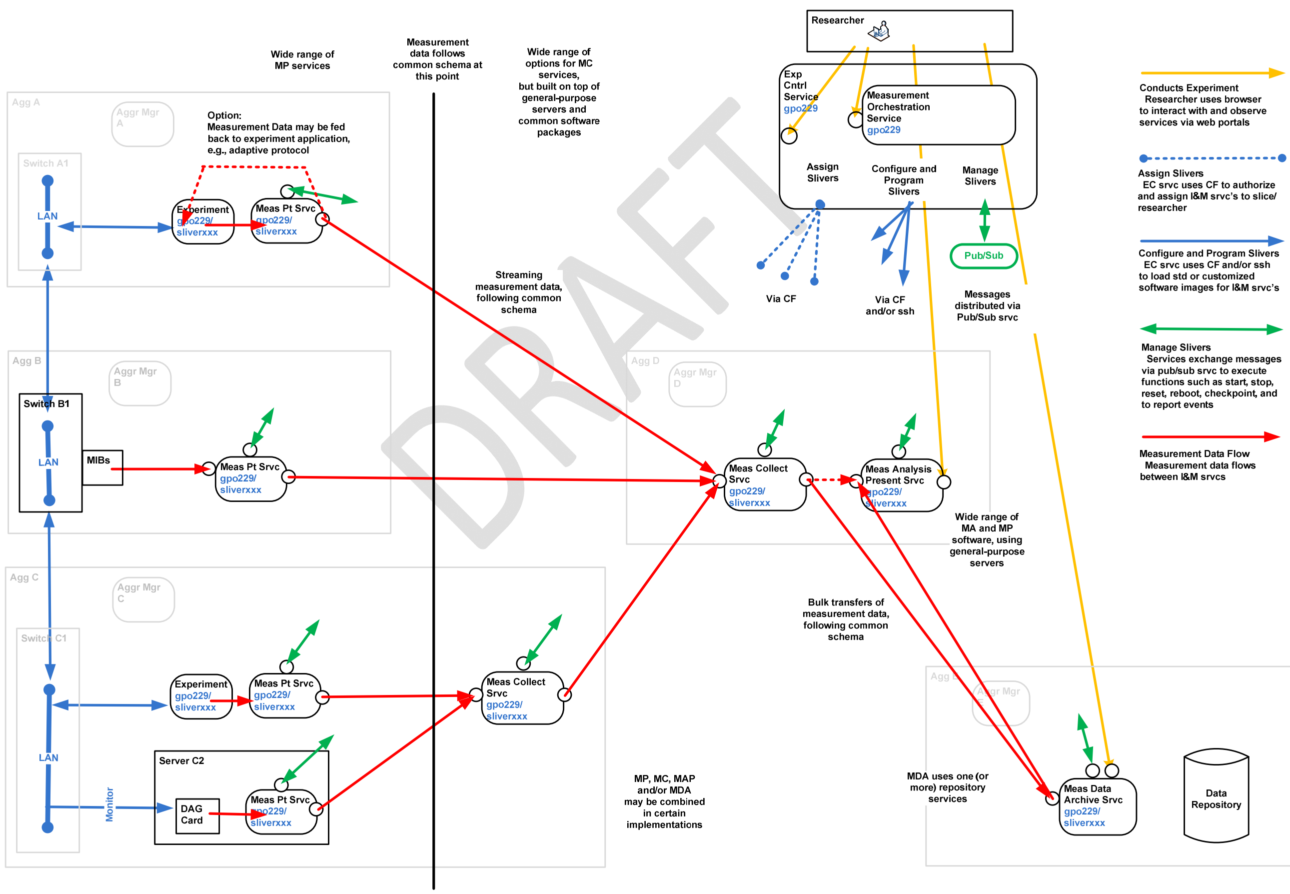

Figure 1-1, Basic Configuration of I&M Services

A collection of I&M services (resources), which have been assigned to a researcher/slice for an experiment by the CF, in the same manner as other resources are assigned to a researcher/slice for an experiment.

Measurement Orchestration (MO) service: p/o Experiment Control (EC) service, typically utilizes a scripting language to orchestrate I&M services.

Measurement Point (MP) services: instrumentation that taps into a network and/or systems, links and/or nodes, to capture measurement data and format it using a standardized schema; wide variety, but all output measurement data adheres to a common schema; measurement data typically includes time stamps.

Measurement Collection (MC) services: programmable systems that collect, combine, transform and cache measurement data.

Measurement Analysis and Presentation (MAP) services: programmable systems that analyze and then present measurement data.

Measurement Data Archive (MDA) services: measurement data repository, with index and portal.

3.3 Basic Measurement Data Flows

Typical flow of measurement data in basic use case: from MPs, through MCs, to MAPs.

Intermediate and final files stored in MDA.

Extended use case (under study by IMF project, Dutta, 1718): take some MD, and feed back to experiment. (IMF project is studying adaptive protocols)

3.4 Range of I&M Service Implementations

Expect range of implementations from small-scale to large-scale. Must be able to build large-scale implementations.

Small-scale implementations might put all I&M services within one aggregate, and even in one server.

Large-scale implementations might have I&M services distributed over many aggregates, with MD flowing between services, and orchestration mechanisms based upon message exchanges between the distributed services and orchestration tools included within the experiment control tools.

4 Interfaces, Protocols and Schema for Measurement Data (MD)

Interfaces between measurement services

Protocols for MD flows

Schema for MD

4.1 Interfaces and Protocols for MD Flows

Standardized interfaces between measurement services

A measurement service may implement one or more standardized interfaces

Interfaces must be bound together when measurement services are assigned to a researcher/slice.

MD flows are transported between measurement services using GENI Measurement Plane (see Section 8)

Classifications:

Pt-to-pt vs pt-to-multipoint

Push vs pull

Stream vs bulk transfer

Disconnected operation expected, or not.

Streaming data

One case being defined in IMF project

Streaming protocols?

Flow control?

Setup, verify, start, pause, stop, repair

Bulk-transfer of data

Access protocols to pull data: snmp; ssh

Web services (SOAP, WSDL)

File transfer protocols: scp; ftp

Setup, verify, start, pause, stop, repair

4.2 Schema for MD

Suggested at GEC6 meeting: Common schema for MD.

Goal: MD follows common schema after it leaves the MP service

Issue: Is there only one common schema, or are at least a limited number of types?

Consider OGF schema being developed in OGF NM WG (Swany, Univ Delaware, 1788)

See IMF project (Dutta, NC State, 1718)

See Data-intensive cloud project, using Amazon S3 resources (Zink and Cecchet, UMass Amherst, 1709)

See recent DatCat project at http://www.datcat.org/ that provides a catalog that indexes Internet measurement data, which you can then find, annotate, and cite. (Klaffy, DAIDA)

See Crawdad project at http://crawdad.cs.dartmouth.edu/ (Kotz, Dartmouth)

Schema includes:

File or stream identifier

Annotation, or meta data

Measurement data

File or stream identifier:

Data record identifier

Slice, owner, encryption

Also experiment, sub-slice, and/or run identifiers, as provided by researcher

Access rules, privacy

Issue: How is this done in a streaming situation?

Annotation, or meta data:

Source, e.g., specifying a monitor link on a VLAN; requires identifiers for nodes and links

Anonymization?

Lifetime?

Additional processing functions applied, by which service, when, etc.

Provenance record?

Type of measurement data:

Type 1: Data with timestamps

Type 2: Logs or events with timestamps

Type 3: Memory or table dumps

Measurement data, type 1:

Known mappings

Data values, timestamps

Error estimates

Filters applied

Start, stop triggers; start, stop times, without or with gaps

Specified triggers and actions, recorded within data.

Measurement data, type 2:

Measurement data, type 3:

5 Ownership of MD and Privacy of Owners

5.1 Ownership of MD

When MD is collected by a particular researcher/slice for an experiment, that MD is "owned" by that researcher/slice, and they may or may not wish to share it with others.

Such MD must only be accessible to authorized users, as defined by the researcher/slice, and/or the acceptable use agreements they have 'signed".

For example, measurement data may (or may not) be visible to the PI associated with the slice.

For example, measurement data may (or may not) be visible to the operations staff.

Thus, different views of the same data must be provided depending on authorization level.

Authorization must be enforced by the CF, and should be resilient to attack.

For further security, the owner of MD may wish to encrypt the data.

5.2 Privacy of Researchers/Slices

Since GENI resources are typically sliced and thus shared with multiple researchers/slices, some MD will contain information for multiple researchers/slices, directly or indirectly.

When such MD is provided to a particular researcher/slice, it must be filtered in such a way as to ensure the privacy of the other researchers/slices, i.e., one researcher must not be able to see, or infer, MD from another researcher/slice.

5.3 Privacy of Application Users

In some cases, MD will contain information on an application being run in GENI, and thus on its users.

The application may be a networking application, i.e.,an ISP running within GENI.

The application may be a service, i.e., a social networking site.

In many cases, the privacy of the users is protected by law, and must not be compromised.

One strategy is to anonymize the MD as it is collected, although it has been shown that this is difficult to implement.

The simplest strategy is to avoid collecting any MD that could compromise privacy.

6 Interfaces, Protocols and API's for Using I&M Services

6.1 Assign I&M Services

Discovery, authorization and assignment of I&M services (resources) to a researcher/slice, done via GENI CF interfaces, protocols and APIs. All traffic is carried in the GENI Control Plane.

Expect many I&M services to be based on a standard configuration, with a sliver dedicated to the researcher, which can then be customized

After authorization, CF assigns I&M sliver to researcher/slice, and then configures it, i.e., installs keys, so that the researcher can access it to configure, program and manage it.

6.2 Configure and Program I&M Services

Researcher via Experiment Control service (tools) loads standard or customized software images for experiment.

Uses CF and/or ssh into target sliver, with keys loaded during assignment. All traffic is carried in the GENI Control Plane.

6.3 Manage I&M Services

Researcher via Experiment Control service (tools), including MO service, manages the setup and running of I&M services

Protocols for researcher/experiment control tools to access APIs, include:

Xml-rpc

web services (SOAP, WSDL)

APIs for setting up and running I&M services

APIs for MP services

APIs for MC services

APIs for MAP services

APIs for MDA service

All traffic is carried in the GENI Control Plane.

Interfaces between researcher/experiment control tools (including MO service) and I&M services for coordination of services

Messages to coordinate services and report events

Functions such as start, stop, reset, reboot, checkpoint.

As done in OMF/OML, via "Control VLAN"

Content of messages?

Protocol of messages?

Use of Pub/Sub to route messages, as done in OMF/OML using pub/sub service based on XMPP?

All traffic is carried in the GENI Control Plane.

Issue: Should a common message exchange service be established in the GENI Control Plane, such as a pub/sub service based on XMPP?

6.4 Observe I&M Services

Researcher observes I&M services , e.g., checks status or views MD

Protocols for researchers (also administrators and operators) to observe running services, etc., including:

web GUI

7 Basic GENI I&M Use Cases

7.1 For Researcher

To support experiments

GENI project examples:

Instrumentation Tools (INSTOOLS) project (Griffeon, Univ Kentucky, 1642), provides standardized I&M services for students

OMF/OML project (Ott and Gruteser, NICTA and Rutgers WINLAB, 1660), provides easily programmed I&M services for wireless testbed, that can include disconnected components

7.2 For Operator

To support operations of GENI infrastructure

GENI-related examples:

perfSONAR project (Swany, Univ Delaware, 1788), monitors Internet2 backbone and other parts of ProtoGENI infrastructure

8 MD Transport via GENI Measurement Plane

Currently, the GENI infrastructure has defined a Control Plane and an Experiment Plane.

This separation of Control and Experiment Planes is intended to prevent experiment traffic from affecting control traffic, no matter what the nature of the experiment traffic.

This separation of Control and Experiment Planes is consistent with substrate implementations where servers have two NICs: one for Control (IP) and one for Experiments (Layer 2, typically VLANs). For example, this is the approach used in ProtoGENI.

To transport MD traffic between I&M services, a GENI Measurement Plane must be defined.

8.1 GENI Control Plane

The GENI Control Plane uses an IP network, carried via a GENI backbone implemented on I2 and/or NLR, with public IP addresses, that should be reachable from the Internet.

It is expected that a GENI researcher can connect into the Control Plane via the Internet.

Each GENI site is expected to have a good connection to the backbone, without any significant bottlenecks, and the backbone is expected to have high capacity and low latency.

Control traffic to and from each site to the backbone is expected to be given the highest priority, so that no other traffic can affect it, e.g., a large flow of experiment traffic cannot block the control traffic required to manage an experiment.

8.2 GENI Experiment Plane

The GENI Experiment Plane uses a switched Layer 2 network, with Layer 2 connections switched within and between aggregates, and carried via a Layer 2 GENI backbone implemented on I2 and/or NLR.

The use of Layer 2 connections allows experiments to use a Layer 3 protocol that is not IP.

In many cases, the Layer 2 connections will carry a multi-point VLAN dedicated to a slice/experiment, interconnecting the aggregates/components assigned to that slice/experiment.

The use of a VLAN to interconnect the aggregates/components assigned to a slice/experiment, permits access to aggregates implemented with private IP addresses (assuming that the private IP addresses in multiple aggregates do not overlap).

Furthermore, a slice/experiment will have access to the Internet via a VLAN, to allow for various “at scale” experiments.

Each GENI site is expected to have a good connection to the backbone, without any significant bottlenecks, and the backbone is expected to have high capacity and low latency.

Experiment traffic to and from each site to the backbone is expected to be given a lower priority than control traffic, so that experiment traffic may be affected (disrupted) by control traffic.

Furthermore, it is possible that experiment traffic from one experiment may affect (disrupt) experiment traffic from another experiment.

Management of the QoS for traffic from multiple experiments is still under study.

8.3 GENI Measurement Plane

The MD traffic flows are expected to include both incremental blocks of data (streams), and occasional large file transfers (bulk transfers).

MD traffic flows (streams or bulk-transfers) will be transported between I&M services by a defined GENI Measurement Plane.

How will this Measurement Plane be implemented?

Option 1: Carry all MD traffic flows using a dedicated measurement VLAN.

As is done in OMF, where a measurement VLAN is defined (called "Control") for this traffic, separate from "Management" (equivalent to GENI "control") and "Experiment" VLANs.

However, this requires the substrate nodes/servers to have three NICs, which is a significant complication.

Option 2: Carry all MD traffic flows using the same IP network that supports the Control Plane.

Where an I&M service, e.g., MP, MO or MAP, is built on a server with two NICs, these flows would use the NIC normally used for control traffic.

The MD traffic flows are expected to include both incremental blocks of data (streams), and occasional large file transfers (bulk transfers).

If the MD traffic flow rates are not too high, this should not burden the access from an aggregate/server to the Control Plane.

How can the rate be set and controlled to carry the necessary MD traffic and still avoid disrupting the control traffic?

Furthermore, large bulk transfers must be broken up, so as not to burden the access from an aggregate/server to the Control Plane.

How can this be done?

Assume: Once the traffic reaches the GENI backbone, it can be carried without affecting other traffic.

Option 3: Carry most MD traffic flows using the same IP network that supports the Control Plane, but for high-rate MD traffic flows, define a dedicated measurement VLAN for the slice/experiment, using the same Layer 2 network that supports the Experiment Plane.

Where an I&M service, e.g., MP, MO or MAP, is built on a server with two NICs, these flows would use the NIC normally used for experiment traffic, and would follow a VLAN to the other services.

It is possible that the measurement traffic could affect (disrupt) experiment traffic from another experiment, or vice-versa. Management of QoS would be helpful, but is still under study.

9 Discovery, Authorization, Assignment and Binding of GENI I&M Services

Discovery, authorization, assignment and binding of resources to a slice is done from Experiment Control service (tools), via GENI CF interfaces, protocols and APIs

In a similar manner, discovery, authorization, assignment and binding of GENI I&M services to a slice, is done from Experiment Control service (tools) (including Measurement Orchestration service), via GENI CF interfaces, protocols and APIs

Expect many I&M services to be based on a standard configuration, with a sliver dedicated to the researcher, which can then be customized

How is an I&M service standard configuration maintained and provided?

Provided as a load for a server or VM, that is then assigned to a slice/experiment?

How is this load customized?

How is the customized load stored and recalled?

Preferred: Each aggregate should include servers; VLAN switches; I&M services.

Binding is required to "connect" or "stitch" the I&M services.

This may involve finding a path from one service to another, and setting up a connection between services. If this process cannot be completed, then different services may need to be assigned.

10 Measurement Orchestration (MO) Service

Expected to be part of Experiment Control service (tools)

Goal: Independent of CF

Is this a service within the Experiment Control Service (tools) with a particular API?

What is the nature of the API?

Expected to be driven by a scripting language – to coordinate measurement process

What features are required in language?

What features are desired in language?

Best current example: OML in OMF (Ott and Gruteser, NICTA and WINLAB/Rutgers, 1660)

Consider: workflow description languages (www.gridworkflow.org)

Expected to call APIs for setting up and running I&M services

APIs for MP, MC, MAP and MDA services

Expected to support protocols and messages for coordinating I&M processes , as done in OMF/OML, via Control VLAN

Pub/sub based on XMPP used in OMF/OML (Ott and Gruteser, NICTA and WINLAB/Rutgers, 1660)

11 Measurement Point (MP) Services

11.1 Functions

Includes: instrumentation (sensors) to tap into a network and/or systems, links and/or nodes

Captures measurement data, and provides necessary caching

Filters as necessary, to capture only essential data.

Always an issue: capturing all data often leads to very, very large amounts of measurement data, that is impractical to store and/or transport.

May also map data relevant to a particular slice/experiment, i.e, identify packets from a particular source

Formats data using a standardized schema

Responsible for accurate and complete use of the schema

Responsible for making accurate timestamps

11.2 Types of MP Services

Expect a wide range of MP services

Options include:

MP embedded within experiment code, provides exactly the measurement data desired

MP associated with a testbed resource (e.g, VM server or network switch), that provides measurement data to be shared by many slices/experiments/researchers.

MP associated with a testbed resource (e.g, VM server or network switch), that provides measurement data for a particular slice/experiment/researcher, often using a virtualized resource.

Preferred: MP embedded instrumentation, since data is accurate and is easily identified, but must not be intrusive.

Preferred: When MP includes a specialized sensor, use a general-purpose server to gather and process measurement data.

11.3 Gathering Measurement Data from Instrumentation (Sensors)

Options for gathering measurement data from instrumentation (sensors):

Pull (poll) data using protocol such as SNMP

Pull (poll) data using protocol such as ssh and a client for an API

Receive traps (messages) from an entity

Gather logs or tables from an entity

Receive data via a flow mechanism ?

Receive data via a pub/sub mechanism

11.4 Types of Instrumentation (Sensors)

Options for instrumentation (sensor):

Passive tap, just gathers data

Active probing, injects signals to understand the state of an entity, i.e., transit time through a network

Instrumentation (sensors) that tap into network and systems

Often provide only basic signals, which must somehow be filtered and identified for a particular slice/experiment.

Link sensors - deployed on network links via taps, provide basic link signals

Issue: Require monitor links into VLANs for link sensors.

Preferred: Link sensors could be general purpose servers with measurement software

Preferred: Programmable systems connected to sensors, e.g., DAG appliance, which transforms basic signals into data suitable for more standard analysis, e.g., framing and flow export.

Node sensors - deployed on all systems connected by links

Provide basic utilization/state/configuration data of servers and VM containers

Provides useful forensic data, i.e., dumps

12 Time-Stamping MD

Almost all of the measurement data must be time-stamped in the MP service.
Accurate time stamps are almost always essential

Options for determining time:
Network Timing Protocol (NTP), accuracy within a few ms (fair)
GPS, accuracy within ? (best)
New: IEEE 1588 Precision timing Protocol
Proposal: RAD clock

13 Measurement Collection (MC) Services

Programmable system that collects, combines, transforms and caches measurement data

Expect wide range of options to collect measurement data

Expect wide range of options to combine measurement data

Expect wide range of options to transform measurement data

Preferred: Use general-purpose servers, with collection software

Expect to be based on a standard configuration, with a sliver dedicated to the researcher, which can then be customized

Expect sliver to be deeply programmable by researcher

Standardized configuration exports known APIs

Receives measurement data in standardized schema

Outputs measurement data in standardized schema

Typically includes cache for short term storage capability

May include a round-robin cache, to limit overall size

Uses MDA service for long-term storage

14 Measurement Analysis and Presentation (MAP) Services

Programmable systems that analyze and present measurement data

Expect wide range of options to analyze measurement data

Expect wide range of options to present measurement data

Preferred: Use wide range of available analysis and presentation tools

Preferred: Use general-purpose servers, with analysis and presentation software

Expect to be based on a standard configuration, with a sliver dedicated to the researcher, which can then be customized

Expect sliver to be deeply programmable by researcher

Standardized pieces export known APIs

Receives measurement data in standardized schema

Outputs measurement data in standardized schema and/or presentation formats, such as accessible via a web GUI

Uses MDA service for long-term storage

15 Measurement Data Archive (MDA) Services

15.1 Functions

Measurement data repository, with index and portal

Suggested at GEC6 meeting: Common mechanism for accessing data

Need to archive measurement data, I&M service configurations

High capacity data repository deployed for use by GENI slices

How is this allocated to a GENI slice/experiment/researcher?

How are files identified?

What is access method?

How is access control done? authentication? encryption?

Different views of the same data will be available depending on authorization level

User-specified privacy

Which can be made public, i.e., used in a published paper?

Resilient to attack

15.2 Implementation

Preferred implementation: use available system, that can scale

Example: Data-intensive cloud project, using Amazon S3 resources (Zink and Cecchet, UMass Amherst, 1709)

Example: Experiment Mgmt System from DOR project (Lannom and Manepalli, CNRI, 1663)

Example: DatCat project at http://www.datcat.org/ that provides a catalog that indexes Internet measurement data, which you can then find, annotate, and cite. (Klaffy, CAIDA)

Example: Crawdad project at http://crawdad.cs.dartmouth.edu/ (Kotz, Dartmouth)

15.3 Use

Need data catalog to index data in the repository

Need portal to allow researcher to find and examine data

Data catalog (file system) needs to be arranged for each slice/experiment/researcher, to give the view(s) they want

Example: DatCat project at http://www.datcat.org/ that provides a catalog that indexes Internet measurement data, which you can then find, annotate, and cite. (Klaffy, CAIDA)

16 Additional GENI I&M Use Cases

16.1 For Researcher

To support experiments

Suggested at GEC6 meeting: Gather in one document the measurements desired for experiments by other projects, i.e., target problems

What use cases need to be supported by the end of Spiral 2 to support experiments?

16.2 For Operator

To support operations of GENI infrastructure

What use cases need to be supported by the end of Spiral 2 for GMOC? For other GENI operations groups?

17 Glossary

The current GENI glossary can be found at: **http://groups.geni.net/geni/wiki/GeniGlossary**

18 References