GENI Instrumentation and Measurement Systems (GIMS) Specification

GDD-06-12

GENI: Global Environment for Network Innovations

December 4, 2006

Status: Draft (Version 0.4)

Note to the reader: this document is a work in progress and continues to evolve rapidly. Certain aspects of the GENI architecture are not yet addressed at all, and, for those aspects that are addressed here, a number of unresolved issues are identified in the text. Further, due to the active development and editing process, some portions of the document may be logically inconsistent with others.

This document is prepared by the Facility Architecture Working Group.

Editor:

Paul Barford, *University of Wisconsin*

Contributing workgroup members:

Jack Brassil, *HP Labs*

Ted Faber, *USC/ISI*

Jay Lepreau, *University of Utah*

Larry Peterson, *Princeton University*

Steve Schwab, *Sparta*

John Wroclawski, *USC/ISI*

Other contributors:

David Andersen, *Carnegie Mellon University*

Tom Anderson, *University of Washington*

Suman Banerjee, *University of Wisconsin*

David Kotz, *Dartmouth University*

Steve Muir, *Princeton University*

Sanjoy Paul, *Rutgers University*

Robert Ricci, *University of Utah*

Timothy Roscoe, *Intel Research*

Stephen Soltesz, *Princeton University*

This work is supported in part by NSF grants CNS-0540815 and CNS-0631422.

Revision History:

Version	Changes log	Date
v0.3	Original version posted	9/15/06
v0.4	Expanded document overview (1.0) Added discussion of design approach and trade-offs (2.1, 2.2) Added resource description framework and example (3.0, 3.1) Modest updates to sections 4 – 8 Added previously omitted bibliographical entries	12/4/06

Table of Contents

1 Document Overview

This document describes the ability to gather, analyze and store data in the GENI infrastructure. This document represents an idealized, architectural perspective of the GENI measurement systems, toward the objectives of supporting the overall GENI mission and the study of the GENI infrastructure as an artifact.

There are three aspects of measurement that are *not* treated in detail in this document. The first are **active measurements** such as traceroute or ping-style probes. These kinds of measurements are based on user-level tools that typically run on end hosts or other types of network systems with more generalized capability, and as such, fall outside of instrumentation environment describe in this document (although new hooks may be developed in the GENI measurement component that these kinds of tools can utilize). The second aspect of measurement not treated in this document is the data required to **secure, manage and operate** the GENI facility. While it is anticipated that the measurement capabilities described below will likely cover a large portion of the security monitoring, management and operation tasks, the focus of this document is on measurement capability required to support experiments in the infrastructure. As this document evolves, it will be extended to accommodate measurement requirements for operations. The third aspect of measurement not treated in this document is **"application layer" logging**. It is assumed that application layer researchers will develop their logging capability that will run within it's slice on a set of components (*e.g.*, a web server log). However, application logging may benefit by measurement hooks available from GENI systems as specified in this document.

The remainder of this document is organized as follows. Section 2 provides an overview of the measurement infrastructure requirements and design, and includes discussions of design approach and trade-offs. Section 3 describes a general framework for defining GENI resources with respect to the basic data types that can be measured and how this data will be gathered. Section 4 describes the design elements of the physical instrumentation that will be deployed in GENI in support of measurement. Section 5 describes the synthesis component of the measurement systems. The design of the archival, analysis and data management systems is described in Section 6. In Section 7, security and access control for measurement systems and data is described. In Section 8, a simple reference design example is given along with a description of how measurement capabilities are accessed from the GENI Management Core [GDD-06-11].

2 Design Overview

The primary objective of the design and development of the GENI instrumentation and measurement systems (GIMS) is to ensure that they are sufficient to support a broad spectrum of empirical network research. The general requirements of GIMS include:

Ubiquitous deployment,

No (or at least measurable) impact on experiments,

Extensibility (*i.e.*, the ability to add new instrumentation and/or new measurement synthesis capability), High availability (at least as available as GENI systems on which experiments are conducted),

Large capacity (*i.e.*, the ability to support a diverse set of simultaneous activities from a large number of experiments),

The ability to measure detailed activity with high accuracy and precision from physical layer through application layer (includes the ability to calibrate measurements),

The ability to specify required measurements for an experiment in a slice (using either standard measurement types from a library or defining user specific measurements) and then having these measurements initialized in the infrastructure when an experiment is activated,

Access control (*i.e.*, the ability to specify what data is available from a particular device or collection of devices, to whom, and for how long),

A large, secure central repository in which collected data can be anonymized and made available to users,

A "data analysis warehouse" where tools for visualizing, interpreting and reporting measurement data can be developed and made openly available.

2.1 Design Approach

The objective of the GENI facility – to provide a large-scale environment for research on network architectures, services and applications – includes support for disruptive research across all layers of the traditional protocol stack. The implication of this objective vis-à-vis measurement capability is that both known data primitives (*e.g.*, 802.3 Ethernet packets or BGP updates or queue occupancy or the routing table on a node, etc.) and data primitives that are yet to be conceived will need to be measured. There are likely to be a number of possible approaches for developing a specification for GIMS with these capabilities and the other capabilities in the requirements list above.

The approach taken in this document is *resource centric*. It begins by specifying a general framework for describing GENI resources, which includes their basic data types and how the resources/data types will be accessed. A key aspect of this process is to be as comprehensive as possible in definition of the basic data types in order to accommodate primitives at higher levels that are yet to be determined. In contrast, the methods for accessing data types should be common, to the extent possible, to all resources. That is, there should be a standard but extensible way for (authorized) gathering of data types in the infrastructure. This framework will be refined by developing a set of reference specifications for measuring well-known data primitives, which will be included in Section 7 of this document.

A challenge in GIMS design using a resource centric approach is that GENI resources are in the process of being defined. We address this by beginning with two abstract resources, *links* and *nodes* and one special case resource, *time sensors*. From these we organize the GIMS specification around three general concepts, (1) instrumentation, (2) measurement synthesis and (3) analysis and archiving, with security and access control applied to each. Items 1 and 2 are the mechanisms for realizing instances in the resource framework, and can be thought of as a hierarchy with instrumentation at the lowest level referring to the physical taps in GENI systems (*e.g.*, on links or within programmable components); synthesis referring to transformations of the signals provided by the taps into meaningful data (*e.g.*, providing layer 2 framing or flow export, or high level aggregates such as routing configuration information); and analysis and archival at the highest level referring to data evaluation and storage in a common repository for future use. These three components form the intrinsic measurement capability of the infrastructure, which is complemented by additional features and capabilities that will accommodate experimental research on instrumentation and measurement, *i.e.*, deploying and testing *new measurement systems and protocols* in GENI.

2.2 Design Trade-offs: Specialized Systems vs. Intrinsic Capability

Like the other aspects of the GENI infrastructure, the scope, capabilities and rollout of the measurement systems will depend on a variety of factors including facility design, budget, user requirements, management requirements, security requirements, and availability of hardware/software for the purpose of measurement, among other factors. An important design issue that must be considered is how/where measurement capabilities will be realized in the facility. This issue can be boiled down to the question of whether to deploy special-purpose measurement hardware versus enhancing the general-purpose programmable hardware being developed for GENI to support required measurements.

It can be argued that special-purpose measurement systems (*e.g.*, Endace DAG cards) offer several advantages. They are designed and built to operate at line rate, and many such systems are available for very high-speed (*e.g.*, 10Gbps) links today. In this sense, they almost certainly offer a price/performance advantage over the general-purpose systems. By virtue of the fact that the specialized systems are “external”, they do not pose a risk in terms of introducing a load on the general-purpose systems being used in experiments. Special purpose systems are also deployed and in use today, which means that there is already community knowledge on how to use them and an existing analysis tool suite for data they collect. Recreating high-speed measurement capability that exists today on commercial products will be a non-trivial enterprise that may not make sense in the short or medium term. Finally, some special purpose measurement systems have some ability to be programmed, which may enable them to fit more directly into GENI.

Enhancing the programmable general-purpose GENI systems with the required measurement capability has the distinct advantage of making measurement truly intrinsic to the infrastructure. This approach is a “cleaner” solution from a design, deployment and management perspective, and a better fit with the overall GENI mission. The risk, however, is that it may be determined that it is infeasible to support certain kinds of required measurements on the general-purpose hardware (*e.g.*, the hardware/OS cannot do what is required), or that certain common types of measurements consistently cause degraded performance on the general-purpose systems.

It seems likely that the deployed solution will be some kind of hybrid between the two different approaches. The advantages of special-purpose systems would seem to dominate (cost-performance, existing solution) especially in the short and medium term. However, it is anticipated the actual realization will evolve as methods for incorporating measurement capability into general-purpose systems are developed, implemented and deployed.

This document reflects a design that includes both special-purpose measurement systems and measurement built into general-purpose GENI components.

3 GIMS Resource Description Framework

This section provides a general framework for describing the basic data types and access methods associated with resources in GENI. The definition of resource in GIMS is consistent with the definition used in other GENI documents (*e.g.*, [GDD-06-11, GDD-06-15]), *i.e.*, that GENI components consist of a collection of physical and logical resources, and in the GIMS context all of these may be something that could be measured. This framework will be consistent with the notations for resource descriptions being developed in other GENI documents and will extend them where necessary to support the GIMS requirements. This framework will also extend basic resource descriptions – at least by adding access methods (which could include the protocol, timing, sampling methods, storage, etc.).

After defining resource metrics and their access methods, the physical (*e.g.*, taps, sensors, counters) and logical (*e.g.*, synthesis functions that transform data primitives into required metrics) infrastructure required to support these measurements must be determined. As mentioned above, this will depend on a variety of issues. As GIMS evolves, an additional consideration will be whether new measurements will require new infrastructure or can be synthesized with existing infrastructure.

The following are open questions about the GIMS resource description framework:

1) What are the details of the notation and format for the resource description framework (consistent with other GENI documents)?

2) How should the framework be extended to metadata about experiments that will be stored in GIMS repository?

3.1 Resource Description Example

As the specifications and notations for expressing GENI resources evolve, the process of defining the physical and logical components of GIMS can be bootstrapped using two abstract resource types – links and nodes. *Links* are the physical media (*e.g.*, copper, fiber, air) that connect nodes. *Nodes* are programmable components on which slices can be instantiated and experiments conducted. The assumption is that the combination of these resources form an infrastructure that will support experiments from physical through application layer and that measurements will be required throughout. The generality of this example represents a bottom up approach to GIMS, much of which is likely to remain relevant as resource definitions come into focus. No attempt is made at using a systematic notation for this example – it is purely descriptive – and the list of items beneath Links / Nodes represent **metrics that should be measured** by GIMS (**NOTE:** the following are not meant to be complete lists of metrics for either resource):

1) Links:

   a) Basic characteristics:

      i) Media – *e.g.*, single mode optical fiber,

      ii) Location – *e.g.*, between nodes A and B,

      iii) Distance – *e.g.*, 40 miles,

b) Basic signals:

i) Type – *e.g.*, EM pulse from a specific transponder type,

(1) Characteristics – *e.g.*, attenuation,

c) Aggregates:

i) Encodings – *e.g.*, method(s) available for encoding bits,

(1) Synthesis method – *e.g.*, signal pattern for a given encoding,

(2) Capacity – *e.g*, maximum bit rate per second,

(3) Characteristics – *e.g.*, bits transmitted or bit error rate per second,

ii) Framing / Packets – *e.g.*, method(s) available from framing / packetizing,

(1) Type – e.g., 802.3 Ethernet frame or IPv4 packet

(2) Synthesis method – *e.g.*, encoding and signaling associated with framing / packets,

(3) Format – *e.g.*, header fields,

(4) Characteristics – *e.g.*, frames / packets transmitted or throughput or goodput,

d) Access methods:

i) Authentication – *e.g.*, credentials A, B or C are required to access this resource,

ii) Authorization – *e.g.*, credential A allows access to the set of X metrics on this resource,

(1) Required anonymization – *e.g.*, authorized access include anonymization of set of Y metrics on this resource,

iii) Protocol:

(1) Type – *e.g.*, use access protocol Z to gather specified metrics,

(2) Parameters – *e.g.*, take N samples over M minutes with timestamps and transfer results to repository location R.

2) Nodes:

a) Basic characteristics:

i) Location – *e.g.*, Lat. / Lon. or machine room label,

ii) Device type – *e.g.*, name and rev. number of widget,

iii) Device configuration – *e.g.*, name and rev. number of OS,

iv) Device dynamics – *e.g.*, CPU or memory utilization,

b) Basic signals:

i) Type – *e.g.*, EM pulse from a specific transponder type or an 802.3 Ethernet frame,

(1) Characteristics – *e.g.*, attenuation or frames transmitted / received,

c) Aggregates:

- i) Encodings – *e.g.*, method(s) available for encoding bits,
  - (1) Synthesis method – *e.g.*, signal pattern for a given encoding,
  - (2) Capacity – *e.g*, maximum bit rate per second,
  - (3) Characteristics – *e.g.*, bits transmitted or bit error rate per second,
- ii) Framing / Packets – *e.g.*, method(s) available from framing / packetizing,
  - (1) Type – e.g., 802.3 Ethernet frame or IPv4 packet
  - (2) Synthesis method – *e.g.*, encoding and signaling associated with framing / packets,
  - (3) Format – *e.g.*, header fields,
  - (4) Characteristics – *e.g.*, frames / packets transmitted or throughput or goodput,
- iii) Table – *e.g.*, the result of a calculation from a service or application on a node,
  - (1) Type – e.g., a BGP table,
  - (2) Characteristics – *e.g.*, update frequency,

d) Access method:

- i) Authentication – *e.g.*, credentials A, B or C are required to access this resource,
- ii) Authorization – *e.g.*, credential A allows access to the set of X metrics on this resource,
  - (1) Required anonymization – *e.g.*, authorized access include anonymization of set of Y metrics on this resource,
- iii) Protocol:
  - (1) Type – *e.g.*, use access protocol Z to gather specified metrics,
  - (2) Parameters – *e.g.*, take N samples over M minutes with timestamps and transfer results to repository location R.

In the examples given above are there is an intentional duplication of several metrics. As noted in Section 2, the decision on how and where to best measure these duplicates is a matter of carefully considering requirements for users, management and security and weighing the associated design tradeoffs.

4 Instrumentation

The strawman resource descriptions above lead to the first element of the realization of GIMS, which is the set of sensors deployed in the infrastructure. There are three types of sensors: (1) link sensors, (2) node sensors, (3) time sensors. The purpose of the sensors is to provide the basic signals that can be synthesized into a variety of specific, well-formed data that will available as measurements to experiments running in the infrastructure. This design represents one possible instance of GIMS, which will be refined as GENI requirements, resource descriptions and other aspects of the facility design come into focus.

4.1 Link Sensors

Link sensors are deployed on all (or some portion) of the physical links in the infrastructure and enable signals transmitted on those links to be extracted. At this level, no assumptions are made as to what the signals mean—interpreting signals is part of the synthesis activity. Signals on the links can be either light or electrical impulses. In the case of **fiber/light** signals, the standard method for tapping a link is to install a passive 50/50 Y-splitter (*e.g.*, [SPLIT]) that makes a duplicate of the light signal on the primary path available on a secondary path that will terminate at a collection system (described below). Standard optical Test Access Port devices such as those offered by Net Optics, (*e.g.*, [TAPS]) and others are also commonly used to mirror light signals. These devices typically do some kind of standard layer 2 framing such as 802.3 (*i.e.*, some basic synthesis), which will make them applicable to a fairly broad range of measurement activities, especially in the short term. However, in the most general sense of the architecture, these taps are not ideal for GIMS unless they are programmable and enabled flexibility in how framing is defined.

In the case of signals on **copper/electrical** links, there are several standard methods for instrumentation that are not unlike those for fiber/optical links. The first is to use basic port mirroring capability available on many types of network nodes. This approach requires the network node to copy all signals on a given port to a monitoring port. For high-speed links, resource demands in this configuration can be considerable—to the extent that they could alter the performance or behavior of the network node. A second method is the "bump in the road" approach, which refers to physically breaking the link between two nodes through a mirroring device similar to the test access port devices mentioned above. In the case of copper/electrical links, this is likely to be the best approach in the short term although, in terms of off-the-shelf solutions, it suffers from the same problems of imposing a fixed type of layer 2 framing on basic signals.

Ideally, sensors on either the fiber/optical and copper/electrical links will be taps that mirror signals on the links. Like current commercial taps, these will be rack-mount devices that can be installed next to GENI nodes. These devices can provide some level of measurement synthesis such as layer 2 framing for standard types of frames, but should pass through the basic signals and also support programmability for new kinds of framing.

Link sensors for **wireless** environments have different characteristics than wireline link sensors but basic the capabilities in many respects are similar. There is a need to be able to monitor the physical environment with respect to the signals being transmitted between multiple nodes simultaneously. This means that separate RF sensing devices with (perhaps) higher quality antennas will have to be deployed in order to gather a complete picture of transmission and interference patterns in a given environment.

The following are open questions about link sensors in GENI:

1) What are the details of the requirements and capabilities for sensors in wireless environments?

2) How can DWDM links be tapped? This may be able to be done with optical splitters, but there is likely to be a much more substantial burden on the collector system.

4.2 Node Sensors

Node sensors are deployed on all nodes interconnected by links in GENI. The essence of node sensors is that they provide basic utilization, state and configuration information about components and/or subsystems in a node. There are two canonical examples of node sensors. The first is the /proc file system that is available in Linux. Proc is a real time, pseudo file system that provides configuration information and measurements of both underlying hardware components and the processes that are running on the hardware. Examples include CPU speed, cache size, CPU utilization, memory utilization, packets sent, packets received, etc. The second example is the Simple Network Management Protocol (SNMP) Management Information Base (MIB) primitives provided by switches, routers, and other networking hardware. In both cases, the available data is limited by what is provided by the underlying hardware. The following are open questions about nodes sensors in GENI:

1) What is the MIB set for network nodes that will be specifically developed for GENI?
2) What utilization/state/configuration information values will be available from end host nodes deployed in GENI?

4.3 Time Sensors

Time sensors are deployed on all GENI node locations. Time sensors provide a high fidelity, synchronized time source for all GENI nodes that can be used for a variety of purposes. In the GIMS context, time sensors will be used as the basis for applying high precision timestamps to measurements.

The problem of distributed time synchronization has been studied for many years. The Network Time Protocol (NTP) offers the ability to generate timestamps that are synchronized on the order of milliseconds based on using a small number of highly accurate time sources (such as Cesium clocks). The risk with NTP is that it does not provide sufficient precision for a significant set of GENI experiments. In this case, higher fidelity time sources, such as those provided by a GPS, which enable timestamp accuracy on the order of microseconds will be required. GPS receivers can be installed in simple PC hosts that can be rack-mounted in GENI locations. These systems can then provide Pulse per Second (PPS) signals and other data (such as latitude and longitude) to any device in the location. The risk of GPS is that it typically requires an external antenna which can be difficult to deploy in some locations. In this case, software clocks [PV02] or CDMA-based GPS systems are a possibility.

Open questions for time sensors include:

1) What are the mechanisms for applying timestamps to specific measurements for different time sensors available? There are certainly well established methods for existing systems, but it is likely that GENI specific systems will need to provide specific support for this capability.
2) What is the minimum timestamp accuracy requirement for GENI (1ms)?

4.4 Other Instrumentation Issues

There is no specific data buffering or storage requirement for any of the three types of sensors. Ideally, sensors simply provide raw signals, and buffering / short term storage is part of the synthesis component described below. However, it is likely that certain solutions will include some amount of synthesis and buffering. In this case the synthesis component will have to query the sensor using a well-defined protocol (*e.g.*, Simple Common Sensor Interface for PlanetLab [RPKW03]).

5 Synthesis

The second element of the GIMS is the set of systems that collect and synthesize the signals provided by the sensors into meaningful data. Collection and synthesis systems will be physically connected to the sensors available in all GENI locations. While they are referred to in this section as "collection and synthesis systems", their physical instantiation may be a single system or multiple systems. These systems will require specific context from users in order to interpret the arriving signals appropriately. An example of this context, which would be provided during the creation of a GENI experiment, is that light signals from a fiber link (a GENI resource / component available for inclusion in a slice) during a specified period of time, are Ethernet packets. In this case the first aspect of measurement synthesis that is required is layer 2 Ethernet framing.

It is expected that the ability to frame and capture packets will be one of the most important network measurement features in the infrastructure (although GIMS is by no means limited to interpreting signals as packet). There are two examples of collection / synthesis systems that are commonly used today that provide this functionality. The first is a standard PC running the tcpdump utility [TDMP]. These systems rely on standard network interface cards to provide framing. The second are Endace DAG daughter cards installed in a standard PC which generate packet traces using layer 1 signals gathered directly from a network link [DAG]. In each case, the systems can be configured to gather full header / payload traces (or some subset there of), and the host PC can be used to buffer a certain volume of this data before it must be streamed to a larger repository.

The synthesis function can also be more complex. For example, an experiment may require flow-export measurements for specified packet streams. In this case, layer 2 framing along with maintaining appropriate flow records of the observed traffic are required. Another example is a sampling algorithm that saves only certain packets. There is also the possibility of even greater levels of measurement synthesis complexity. For example using collection systems for streaming queries or statistical anomaly detection. In each case, a transformation function (realized as a program) from the basic link signals must be provided by the user or selected from a library of functions that have already been developed. In each of these cases, the requirements of the collection / synthesis nodes is that are programmable thereby allowing users to define exactly what they want to extract from the low level sensors (modulo privileges as explained in Section 6).

From the perspective of node sensors, collection systems will require additional capability. In particular, a data gathering protocol will be required to request transfer of the MIB variables from the node sensors. This data gathering protocol must be light weight, extensible in terms of the kinds of things that can be requested, and include a level of security that is TBD. The most obvious example of an existing protocol in this space is SNMP, although it's well documented deficiencies may preclude it from use in GIMS. Another possibility is the Simple Common Sensor Interface for PlanetLab that has a data query protocol based on HTTP.

The availability of resources (CPU, disk, memory, etc.) on the collection and synthesis systems will obviously limit the extent of these activities. In some cases, this may cause certain experiments to be conducted at times when overall demand is sufficiently low. Regardless of their capacity, collection and synthesis systems are not meant to be anything other than short term repositories for collected data. Therefore, collection systems must have a direct connection with the archival data repository via a secure, transaction oriented (perhaps) protocol. This protocol will be used to move potentially large amounts of data off of the collection systems and into the repository.

It is envisioned that it may be useful in some experiments to have data collection systems at several sites work in close coordination. For example, experiments with collaborative intrusion detection methods may require real-time updates in terms of what to measure and how to measure. Another example are experiments that require simultaneous link and node measurements such as tracing traffic through a system that performs network address translation. In this case, a packet trace from one side of the NAT along with snapshots of the NAT table would suffice. While the signaling required for collaborative monitoring can be implemented at the user level, a secure, real time protocol for coordinating distributed measurement synthesis in GENI is not yet available.

Finally, the task of applying timestamps to measurements is the responsibility of the collection and synthesis systems. Timing sensors will provide input to the collectors which will, in turn, generate and apply the specified timing information to the other measures. Ideally, this capability is available in hardware that can accept GPS timing signals (*e.g.*, Endace DAG cards include a PPS port).

Open questions for collection and synthesis systems and capabilities in GENI include (note that all of these questions may end up becoming issues for GENI services and not issues for the facility architecture):

1) What protocol should be used to facilitate data transfers from link and node sensors to collection / synthesis systems?

2) What protocol should be used to facilitate data transfers from collection / synthesis systems to archive systems

3) What protocol should be used to facilitate coordination between collection / synthesis systems?

4) What is the security model and corresponding mechanisms for the collection / synthesis systems?

6 Data Archive

The third element of the GIMS is the archival repository into which measurements from the collection systems are stored and then made available for analysis. An example of an existing repository similar to what is envisioned for GENI is CRAWDAD, which is used for warehousing and sharing wireless measurements [CRAWDAD]. The archive will have two primary components. The first is the repository itself, which resides on high capacity storage systems that will be located in several GENI sites. The second is a data catalog that indexes the data in the repository. The objectives for the archive are, (1) to provide a secure, reliable repository to GENI users for measurements taken in their experiments, and (2) to provide a standard interface for data extraction so that users can analyze their data at their own sites.

The archival system will interface with the collection and synthesis systems through a data transfer protocol such as the Simple Common Sensor Interface for PlanetLab. This data transfer protocol must be the same as the data transfer protocol used between sensors and collection/synthesis systems for consistency and to enable the possibility of direct streaming into the archival system.

The archive itself will be located in several GENI sites. This enables the archive to be resilient to site-specific failures and will spread the network load for both sensor-to-archive transfers and archive-to-user transfers. A data management system will be used to store and retrieve the measurement data on these systems. An example of a large scale network data management system that has been used for some time is AT&T's Daytona system. Daytona is a file-oriented archive system that provides a SQL-like interface [DAYTONA]. Open questions for the archive include:

1) Details of the physical systems (capacity, processing capability, etc.) on which the archive will run as well as policies and procedures for maintenance and backup of these systems.
2) Details of the of data management system that will run on the physical systems including the interface that it exports to users.

Like the measurements themselves, users creating experiments will have to specify the details of how their data will be archived. This will include the type of data being stored, and the size and lifetime of the storage space required. It may certainly be the case that an individual experiment does not need to use the archive, *e.g.*, an application log. Access to privileges will also have to be specified (this can include access to anonymized data see Section 6)—since data can range from projects in which data is considered private to data that is openly available and permanently archived (*e.g.*, in the case of studies of GENI-wide activity). Details of how users specify their data archive needs and how these are instantiated in GIMS are TBD.

The archival system will also include a separate but related component which is a data index for all data in the archive (such as the Internet Measurement Data Catalog from CAIDA [DATCAT]). This index enables users to find data in the archive. It may also act as a portal pointing to data sets collected in GENI that reside on remote sites.

7 Security and Access Control

There are significant issues of security and privacy for GIMS. The most general requirements for security and privacy are:

1) Sensors, collection/synthesis and archival systems must only be accessible by authorized users,
2) Authorized users must only be allowed access to a specified proportion of a the resources on the measurement systems,
3) Different views of the same data will be required depending on the approved level of authorization.
4) All of the measurement systems themselves must be secured against attack,
5) No aspects of the data that is collected will compromise or violate the privacy of individuals or organizations that source/sink the data.

The mechanisms for addressing requirements #1 - #4 above are all to be determined.

There is a constant tension between desired visibility into measurement data and privacy. However, the privacy requirement (item #5) is very serious since if violated, it could lead to legal action against GENI. Therefore, a set of policies and procedures for requesting and granting access to the entire range of measurements available in GENI will have to be established. In many cases, this will be straightforward. For example, access to CPU utilization on a particular node that is openly available for use by anyone will be immediately authorized. Alternately, access to full packet traces (including payloads) on a link that handles commodity traffic may be possible under any circumstances (HIPPA guidelines will apply).

The mechanisms for addressing privacy issues are to be determined.

8 A GIMS Reference Design

This section provides a reference design and use scenario for GIMS. The model is based on an idealized physical deployment model for GENI that makes simplifying assumptions about architecture. This reference model does not include all envisioned equipment that will be included in GENI (*e.g., wireless* which will be added in coordination with efforts of the wireless working group), and it will evolve to reflect relevant changes in the underlying architecture as it comes into focus. The model also includes examples (not comprehensive) of how a service interface to the underlying GIMS capability might function.

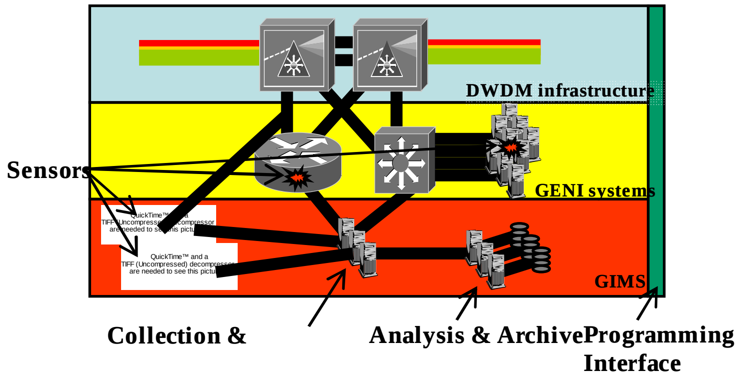
DWDM infrastructure
GENI systems
GIMS
Sensors
QuickTime™ and a TIFF (Uncompressed) decompressor are needed to see this picture.
QuickTime™ and a TIFF (Uncompressed) decompressor are needed to see this picture.
Collection &
Analysis & Archive
Programming Interface

Figure 1. GIMS reference design example at a GENI site

8.1 Physical Deployement

GIMS is based on a physical deployment model consisting of a set of sites interconnected with fiber cables. DWDM signals will be transmitted between the sites. In the most general case, no assumptions about the nature of the signals are made apriori, however for the purpose of this reference model, we assume well defined packets are the standard mode of data transfer. Each site will include rack space, power, physical security, etc. for GIMS equipment.

GIMS equipment at each site, as shown in Figure 1, begins with the physical instrumentation that provides a duplicate of each DWDM signal through a network tap. Dedicated high performance workstations that include DAG measurements cards are connected to each of the taps. These collection systems enable continuous full packet capture, limited synthesis/storage capability and can stream data to additional high performance workstations that enable more robust synthesis. There are also large-scale storage systems in each site into which measurements can be archived. There are also GPS time servers that act as PPS sources for DAG and other measurement timestamps as well as experimental synchronization. Beyond the measurement specific equipment, each site will also contain a variety of systems such as switches, routers (customizable and vendor provided), PCs and other programmable hardware. Each of these pieces of equipment can communicate with collection/synthesis nodes so that data can be gathered per operational and experimental requirements.

At each site, data will pass between instrumentation, collection/synthesis and archive nodes. This communication will take place via a (yet to be defined) data transfer protocol running on each of the nodes. Setting up configurations on GIMS nodes will be also facilitated by this protocol.

8.2 GIMS Configuration

The primary configuration task at a site is specifying the security and access control policies. These policies specify details such as who may have access to which data when, and what level of anonymization is required. Of particular concern is access to packet payload information—an issue to be determined concerns how this information is propagated throughout the infrastructure.

8.3 User Access Services

Each of the sensor/collection node combinations in a co-lo site has a basic set (*i.e.*, all) of measurements that can be made available. Examples include packet traces from links (note, since this reference model only includes packet capture capability, there is only one set of basic measurements available from link sensors), buffer occupancy from routers or CPU utilization from end-hosts. Each node can make this basic measurement capability available through a measurement publish/subscribe system that will be developed as a GENI service. Furthermore, basic measurements can be transformed into new data sets (*e.g.*, transforming packet traces into flows) via synthesis code which can be downloaded and run on synthesis nodes. These transformation capabilities will also be available through the pub/sub service.

When users specify a slice, slivers in that slice can include measurement capabilities and the possible reservation of resources in collection/synthesis and archive nodes. When an experiment is initiated, it will cause basic measurements to flow into collection/synthesis nodes, synthesis transformations to be initiated and measurement data to be placed in archive storage. These activities will be mediated through the measurement subscription service based on the users credentials and the security policies at each co-lo site.

9 References

[GDD-06-11] L. Peterson and J. Wroclawski (Eds), "Overview of the GENI Architecture", *GENI Design Document 06-11*, Facility Architecture Working Group, September 2006.

[GDD-06-15] S. Paul (Ed.), "Requirements for Wireless GENI Management and Control", *GENI Design Document 06-15*, Wireless Working Group, September, 2006.

[CRAWDAD] CRAWDAD, A Community Resource for Archiving Wireless Data at Dartmouth, http://carwdad.cs.dartmouth.edu, 2006.

[DATCAT] CAIDA, The Internet Measurement Data Catalog, http://imdc.datcat.org, 2006.

[DAG] Endace, Network Monitoring Cards, http://www.endace.com/, 2006.

[TAPS] NetOptics, Network Test Access Port Devices, http://www.netoptics.com/, 2006.

[PV02] A. Pasztor and D. Veitch, PC Based Precision Timing without GPS. In Proceedings of ACM SIGMETRICS '02, Marina del Rey, CA, June, 2002.

[DAYTONA] AT&T Labs Research, The Daytona Data Management System, http://www.research.att.com/daytona, 2006.

[RPKW03] T. Roscoe, L. Peterson, S. Karlin, and M. Wawrzoniak. A Simple Common Sensor Interface for PlanetLab. http://www.planet-lab.org/PDN, March, 2003.

[SPLIT] Cisco Systems, WDM Splitter Cable, http://www.cisco.com/, 2006.

[TDMP] tcpdump, http://ftp.ee.lbl.gov/tcpdump.tar, 2006.