Digital Object Registry

Corporation for National Research Initiatives

GENI Quarterly Report – Jan 1, 2010 through Mar 31, 2010

1. Major Accomplishments

The scope of work on this project is to adapt the Handle System and/or the CNRI Digital Object Registry to create a clearinghouse registry for principals, slices, and/or components in at least one GENI Spiral 1 control framework, capable of supporting limited operations in Year 1 and extending that functionality to additional frameworks and to additional functions in Year 2. We will analyze ways in which the Handle System and/or a Digital Object Registry could be used to identify and register GENI software, including experimenter's tools, test images and configurations, and test results. Finally, we will define the operational, scaling, security, and management requirements, plus recommended design approaches, for implementing GENI clearinghouse and software registry services.

During this quarter, we continued participating in various GENI activities and programs, and also continued our collaboration with GENI members including University of Utah and University of Washington members in order to build various services for GENI. We:

- Continued to make available a Digital Object Repository service that enables a unified and consistent mechanism for managing various experimental resources such as test cases, software packages, results, etc.;

- Demonstrated the Digital Object Repository capability to the GPO and GENI members during the seventh GENI Engineering Conference held in Durham, NC;

- Migrated the Digital Object Registry service hosted for the Million Node GENI project to a production environment. This service is capable of managing the operational records of that project;

- Continued to collaborate with the ProtoGENI group to federate the ProtoGENI clearinghouse records into the GENI Federated Clearinghouse;

- Continued to participate in the ProtoGENI biweekly calls.

1.a. Milestones achieved: We are continuing to work on creating an automated process to federate ProtoGENI records into the proposed federated clearinghouse based on the prototype demonstrated in Spiral 1. While the goal is to be able to complete such automated process by the end of March, we faced a few technical challenges that impacted our schedule, the details of which are discussed later in this report.

We have moved the federated service to a production environment, which is a substantial part of a future milestone. We have also moved the Digital Object Registry, which we are hosting for the Million Node GENI project, into a production environment.

1.b. Deliverables: During this quarter, we produced a poster on the various services CNRI is building for the GENI community. We also demonstrated the offered services at the GEC 7. The poster is available on the GENI wiki page for our project.

2. Description of Work Performed

2.a. Activities and Findings During This Quarter

GENI Federated Clearinghouse

The GENI Federated Clearinghouse prototype created during Spiral 1 was presented to GENI members and participants at GEC 7. During this quarter, our goal was to create an automated process that federates ProtoGENI clearinghouse records routinely. This would enable potential clients to view the current state of ProtoGENI managed resources from the federated clearinghouse. As previously reported to the GPO, during the GEC 6, we discussed the possibility of expanding the federated clearinghouse to the PlanetLab and ORBIT clusters. We believe that the federated clearinghouse would benefit the experimenters by providing access to the resource-related information based on a unified service, regardless of the actual location of that information.

Automating the federation process involves:

1. using the ProtoGENI client scripts to interact with the clearinghouse in a programmatic fashion,

2. parsing the information from the ProtoGENI clearinghouse records, and identifying which information changed since the last update,

3. serializing the identified information into 'create', 'update', or 'delete' operations for the federated clearinghouse to process, and

4. finally requesting the federated clearinghouse to process those operations.

The prototype demonstrated during the GEC 6 performed items 2, 3, and 4 above, sequentially and in semi-automated fashion; the first operation was performed manually. Consequently, the core of this milestone entails combining the four steps in an automated fashion. However, we faced certain challenges that affected our schedule. We identified that the ProtoGENI client scripts, which were developed in Python, were not well suited to plug into an external process such as ours, mainly due to the lack of a modular approach in the design of those client scripts. Given that the client scripts were designed mostly from a perspective of some entity setting up a ProtoGENI/Emulab node, where those scripts have a defined role, it is not

surprising that those scripts were not optimized for integrating into a new environment such as ours. However, this resulted in modifying certain ProtoGENI modules; we developed a client-level API that is more generic and widely usable. We also found that some of the modules used by the ProtoGENI client scripts were not Windows compatible; we created a mechanism to address the compatibility issue. Overall, the re-engineered modules not only increased the flexibility of the software for it to be widely usable, but also improved its portability. Once we complete the automated process, we will propose integrating our changes into the ProtoGENI software base.

Acquiring credentials from ProtoGENI to read ProtoGENI clearinghouse records was also challenging. After a few discussions with the ProtoGENI members, we identified that a new credential model was necessary to allow clients like CNRI, who are different from a typical ProtoGENI client. This new credential model would ensure that these atypical clients are not awarded rights to the system beyond what is essential. We believe that integrating security measures such as this is important for any system from operations point of view. ProtoGENI implemented this new credential model soon after GEC 7.

As both of those challenges are now addressed, we would expect the automated federation to be completed in the next quarter. However, we believe the challenges faced and solved, at least partially at this point, have contributed to ProtoGENI and have helped in expanding its usability.

Future Milestones

One of the future milestones in Spiral 2 is to offer the proposed services at a production level. Specifically that entails making the services available 24x7, monitoring the availability of those services, providing services over sufficient network bandwidth, fixing or upgrading the software based on the feedback, etc. In this quarter, we migrated the GENI Federation Clearinghouse service and the Distributed Hashtable Service (hosted for Million Node GENI) into a secure co-location facility. Those services are hosted on a fast Linux server box connected to a 100 Mbps network pipe. An external monitoring service, SiteScope, is configured to check and alert the availability of those services every few minutes. The data from the machine is backed-up daily.

Digital Object Registry Services for Million Node GENI

In the last quarter of 2009 we initiated a Digital Object Registry instance specifically configured for managing the operational records for the Million Node GENI project. During this past quarter we moved that service over to production status. The service allows registering and looking-up information related to node availability. The hosted Registry instance can be thought of as a distributed hash-table where node-related information is hashed based on user public keys. Lookups on the node information can be requested by providing the users' public keys. The Million Node GENI project members have tested the deployed Registry instance. A few bugs

identified during the test run have been addressed. Million Node GENI has integrated the offered Registry instance into their workflow. In order to guarantee higher resilience and greater robustness, we migrated the service to a production environment.

While Million Node GENI is the only project using the offered distributed hash table service, it would be possible to host and provide similar capabilities to other GENI participants. A possible outcome of such an extended collaboration with other GENI participants would be the development of robust and reliable distributed data structures including Lists and Sets. Furthermore, this service could be hosted as an **experiment** on the networking resources provided by the GENI community, as opposed to or in addition to its current status as hosted on standard production Internet facilities. We will examine this possibility for future work.

2.b. Project Participants

CNRI has discussed its project activities with a number of other GENI participants, but all work done this quarter was done by CNRI alone or with the cooperation of ProtoGENI and Million Node GENI personnel. Names and email addresses of CNRI participants are available on the GENI wiki page for the project. Robert Ricci, Leigh Stoller, and Srikanth Chikkulapelly from ProtoGENI and Justin Cappos from Million Node GENI collaborated with us during this quarter.

2.c. Publications

No publications were produced this quarter. CNRI demonstrated the proposed services and displayed a poster during the seventh GENI Engineering Conference. Those documents are available on the GENI wiki page for the project.

2.d. Outreach Activities

CNRI, specifically project PI Laurence Lannom, Giridhar Manepalli and Christophe Blanchi, attended the technical discussions in the ProtoGENI bi-weekly tele-conference and also participated in various GENI mailing lists.