Project 1706 - Plan for making provenance information available to experimenters

NetKarma-S3.b

February 20, 2010

Making provenance available to experimenters: We currently ask the community to donate their GUSH log files. To this end, we prepared a log file submission tool available at http://pti.iu.edu/d2i/provenance/submit-gush-log. We parse the log files and drop the provenance events into the NetKarma Provenance Repository, a persistent Web Service that resides on a server in the GENI Meta-Operations Center (GMOC), located at Indiana University. Contributors can retrieve their provenance from the repository. The repository has a WSDL access API so provenance can be retrieved programmatically. The persistent Axis2 web service is available at the following URL:

http://netkarma.testlab.grnoc.iu.edu:8080/axis2/services/KarmaService

Donators of their GUSH log files will get a message back with information that explains how to query the web service for the provenance graph of their GUSH run. In turn, this will enable GENI experimenters show such provenance information using the CytoScape http://www.cytoscape.org/ visualization tool and Google Earth View.

We identify other sources of provenance information such as Raven provisioning service and GMOC database. We are currently working on harvesting provenance data from these identified sources. We work on processing logs from Raven to obtain provenance information. This new capability will enable GENI experimenters capture provenance such as experiment node locations, time of deployment of software packages and versions of software deployed. We also work on gathering information from the GMOC database. By harvesting information about hosts and circuits from the GMOC database, we plan on augmenting the host descriptions we have in netKarma that are associated with experimental runs. This will help the GENI experimenters relate the provenance on-the-fly to network measurements. We use a rule based solution to parse log files and harvest information from provenance sources. This capability will enable users to specify what types of provenance to be collected at different provenance sources.

We have made a small change to GENI GUSH to make provenance data easier to collect, and are passing the changes back to the GUSH team. We are working on an additional change to GUSH that provides an option to turn on provenance collection in GUSH. In this way, provenance information associated with an experiment run by GUSH will automatically be recorded in netKarma repository and be accessible via netKarma Query API for browsing and visualization. Provenance data will live in Karma repository with unique identifiers (e.g. URLs) pointing the location of associated experiment data.

Support for GENI security will be added to the netKarma repository so that a GENI certificate will be required to query provenance data.