Visualization of Provenance captured by NetKarma

Karma project website: *http://pti.iu.edu/d2i/provenance_karma*
NetKarma project website: *http://pti.iu.edu/d2i/provenance_netkarma*

NetKarma is a tool and repository for capturing and accessing the workflow of Global Environments for Network Innovations (GENI) experiments at multiple layers in the PlanetLab stack, including slice creation, topology of the slice, operational status, and links to measurement data. In GEC-10, we focus on visualizing provenance information for data collected from a GENI experiment using the CytoScape (http://www.cytoscape.org) and Google Earth View visualization tools. We demonstrate this by visualizing the log files generated from Twister (*http://www.iterativemapreduce.org*), a parallel, iterative version of MapReduce, and executing a breadth-first search through a random graph. The application is executed using GUSH (*http://gush.cs.williams.edu/trac/gush*), provenance gathered and visualized using the CytoScape and Google Earth View.

For this demonstration, we identified the GMOC database as an additional source of provenance information and harvest provenance data about hosts. This capability enables GENI experiments to additionally represent provenance such as experiment node locations, operational status and device name. By harvesting information about hosts and circuits from the GMOC database, we augment the host descriptions of experimental runs held in the netKarma persistent provenance service. This additional provenance will help the GENI experimenters relate the provenance on-the-fly to network measurements. As part of satisfying a GEC-10 deliverable, as our ongoing work we will work with providers of instrument traces to tie provenance data in netKarma to relevant network instrument traces. Through this connection, experimenters can access instrument traces by way of netKarma, giving an experiment focused access path. Additionally, we hope through collaboration with the Network Instrumentation & Measurement group to augment network measurement data resident in a measurement repository with provenance.

How will visualization of provenance be of help to GENI experimenters? The idea is that the GENI experimenter loads their run data into NetKarma Provenance Repository; can retrieve their experiment provenance data as a graph in XML format; and can visualize their experiment showing exactly what is going on in the experiment as well as effects that processes have on the backbone. To make the provenance available to experimenters for browsing and visualization, we are asking 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 and Google Earth View tools.