NetKarma
Quarterly Report for Period GEC9-GEC10

Beth Plale – Principal Investigator
Chris Small – Co-Principal Investigator

Summary

- *Submitted changes to GUSH team to make provenance data easier to collect*
- *Satisfied:*
  - *establish plans for making provenance data available to experimenters*
  - *Demonstration at GEC10*
- *Looking ahead:*
  - *identify network instrument traces as a new source of provenance information*

Major Accomplishments

Milestones Achieved

S3.b Plan for making provenance information available to experimenters: Delivered a document with plan for how provenance information will be provided to experimenters.

S3.c Demonstration and Outreach at GEC10: Demonstrated a GENI experiment and display of provenance information for data collected by this experiment. The demonstration included at least one new source, the GMOC database, of provenance information. The team received feedback at the GEC conferences on the kinds of provenance that might be useful. Based on such feedback, the team identified one other source, the network instrument traces, of provenance information to harvest provenance.

Deliverables Made

As per our milestones, the netKarma team delivered a document focusing on our plans for making provenance information available to experimenters and netKarma software suite that successfully was used for a live demonstration in GEC10.

Document on plans for making provenance information available to experimenters: With this deliverable, the netKarma team asked the community to donate their GUSH log files. To this end, the team prepared a log file submission tool available at the

netKarma project website. The log files are parsed and the provenance events are dropped into a public-accessible NetKarma Provenance Repository. Contributors can retrieve their provenance from the repository. The repository has a WSDL access API so provenance can be retrieved programmatically. 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 display provenance information using visualization tools such as CytoScape and Google Earth View.

Software for live demonstration at GEG10: The netKarma team delivered the latest version of netKarma software suite available at project website. Through using this software, during GEC10 the netKarma team demonstrated the visualization of provenance information collected by a GENI experiment running in a PlanetLab slice utilizing a set of nodes. The netKarma team demonstrated all stages of provenance lifecycle using netKarma software suite. These stages include execution of a GENI experiment, provenance capture from the experiment using a rule-based adaptor, storage in the netKarma repository, retrival through the XML query API and visualization using extensions (plug-ins) to CytoScape visualization tool and Google Earth View. For this demonstration, the netKarma team 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.

Description of Work Performed During Last Quarter

Activities and Findings

1. Outreach at GEC10: Engaged with Matt Bishop of UC Davis and Jeffrey Hunker on attribution and how provenance management tools can fit into management of attribution scenarios to aid capture, storage, and retrieval of attribution.

2. Demonstration for GEC10: Successful live demonstration of provenance visualization. The live demonstration showed all stages of provenance lifecycle including provenance capture using a rule-based adaptor, storage in the netKarma repository, retrieval through the XML query API and visualization through extended version of CytoScape visualization tool.

3. Making provenance available to experimenters: We currently ask the community to donate their GUSH log files. To this end, we have a log file submission tool set up at the netKarma web site. We parse the log files and drop the provenance 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. 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. 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.

4. Changes to GUSH: We have made changes to GENI GUSH to make provenance data easier to collect such as by adding timestamps to the logs. These changes are passed back to the GUSH team.

5. Improvements in netKarma: We have improved the code for netKarma Adaptor so it can capture timestamps of provenance events. This allows better provenance and uses such as the creation of movies of provenance history.

6. Identifying other provenance sources: 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.

Project Participants

During this time, key participants in the NetKarma project included: Beth Plale, PI Chris Small, Co-PI Mehmet Aktas, Postdoctoral Fellow, Devarshi Goshal, PhD student, Peng Chen, PhD student, You-Wei Cheah, PhD student, David Ripley, Technical Staff, Robert Ping, Project and Information Management

Publications & Documents

Beth Plale, Bin Cao, Mehmet Aktas, Provenance Collection of Unmanaged Workflows with Karma, journal manuscript under review, 2011.

Visualization of Provenance captured by NetKarma, Demonstration handout at the 6th GENI Engineering Conference http://groups.geni.net/geni/attachment/wiki /netKarma-/GEC-10%20Demo%20Handout.pdf

Outreach Activities

In GEC10, we engaged with Matt Bishop and Jeffrey Hunker about attribution and how provenance management tools can fit into management of attribution scenarios to aid capture, storage, and retrieval of attribution.

Collaborations

We anticipate through collaboration with the Network Instrumentation & Measurement group to augment network measurement data resident in a measurement repository with provenance. We anticipate follow-on discussions on attribution to occur as well.

Planned Activities for next Quarter

As part of 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.

A demo of netKarma with the new identified provenance information sources is planned for GEC11 in July 2011.