NetKarma
Quarterly Report for Period Ending GEC12

Beth Plale, School of Informatics and Computing, Indiana University
Chris Small, InCNTRE, Indiana University

Summary

For the quarter ending with GEC12, we released version 2.5 of NetKarma which enhanced the NetKarma Adaptor and NetKarma visualization plug-ins for Cytoscape to provide more detailed provenance data from the GUSH logs of experiments run on GENI slices. In addition to extracting more detailed provenance from GUSH logs, version 2.5 of the NetKarma plug-in includes richer annotations of the provenance visualized and is more configurable by the user - allowing greater flexibility in visualizing and comprehending the provenance of an experiment. Version 2.5 also includes the NetKarma GMOC Adaptor which automates the annotation of NetKarma provenance graphs with network data from GMOC – this builds on the earlier effort presented in a poster at GEC11 where we annotated provenance graphs with measurements from GMOC.

The NetKarma Provenance System and the NetKarma GMOC Adaptor are installed and running as a persistent service on servers at Indiana University to allow GENI users to process a GUSH log through the NetKarma Adaptor and visualize the resulting provenance graph, including the automated GMOC annotations, without having to install the NetKarma server.

We have continued to engage with the Measurement and Instrumentation team led by Harry Mussman and presented a lightening talk on provenance and initial thoughts on the MDOD at the I&M session at GEC12. We began working with a GENI funded experiment on demonstration of provenance usefulness.

Quarter Deliverables Discussed

S3.f Deliver software and documentation: Version2.5 of NetKarma was released and made available as a service that GENI experimenters could use to upload their GUSH logs and view a provenance graph. Version 2.5 of the NetKarma adapter ingests a log file and can be easily configured to upload the generated provenance notifications to a server running on the GRNOC servers at Indiana University. Using version 2.5 of the NetKarma plug-ins for Cytoscape, experimenters can download and view a provenance graph for their experiment based on the provenance graph ID generated by the NetKarma adaptor when they uploaded their log file. Using this version of the NetKarma adaptor and visualization tools, experimenters do not need to install a NetKarma server themselves.

Features that were added to NetKarma in version 2.5 include:
- Enhancements to the NetKarma Adaptor used to process GUSH logs and ingest the provenance of GENI experiments run on GUSH into the NetKarma Provenance System.

- Enhancements to the NetKarma Cytoscape plug-in for NetKarma that includes more detailed provenance and the addition of annotations to processes, artifacts, and the edges that connect the processes and artifacts in the OPM graphs generated using NetKarma. This

version of the visualization plug-in also has increased configurability by the user as to the details displayed when visualizing provenance graphs.

- The new NetKarma GMOC Adaptor is being released which extracts additional metadata from the GMOC database and annotates the GUSH provenance graphs with data from the GMOC database. This Adaptor is configured to run as a background process and automatically annotates the provenance ingested into NetKarma without any additional steps required by GENI users.

S4.a Provence System Evaluation – Written plan for Evaluation: Our plan for evaluating the NetKarma system is through using GENI experiments to demonstrate functionality. We are working with a GENI funded WiMAX DDOS experiment from Clemson University. This particular experiment came on recommendation from Mark Berman during GEC11. We discussed the project with Richard Brooks, PI from Clemson and plan to present a demonstration of results at GEC 13. The written plan for evaluating NetKarma is available at: http://groups.geni.net/geni/attachment/wiki/netKarma/GENI%20-%20Provenance%20System%20Evaluation.pdf

S4.a I&M Contributions: The GENI Instrumentation and Measurement (I&M) community Measurement Data Object Descriptor (MDOD) contains metadata describing a set of measurement data generated using I&M tools.   The goal of the MDOD is to describe measurement data with the metadata needed to enable experimenters to discover and retrieve from an archive the measurement data of interest to them. Building on a draft version of the MDOD presented at GEC11 and version 1.0 of the GENI Instrumentation and Measurement Architecture documentation, we proposed a high-level schema for the MDOD that reorganizes the metadata content and provides a structure for grouping measurement data descriptors into collections. The goal of the schema outlined in our whitepaper is to enable the capture of provenance regarding the data objects captured in the MDOD as well as provenance of the MDOD itself, while enabling the extensibility to describe new measure data sources and derived data objects as they are developed. The whitepaper, "Provenance Metadata and Extensibility of Metadata Describing Measurement Data", is available at the following URL on the NetKarma wiki: http://groups.geni.net/geni/attachment/wiki/netKarma/MDOD%20Provenance%20and%20Extensibility.pdf

Description of Work Performed Prior Quarter

Version 2.5 of NetKarma was released and made available as a service from the GRNOC servers located at Indiana University. Building on earlier versions of the NetKarma Adaptor and NetKarma visualization plug-ins for Cytoscape, this version provides more detailed provenance data from the GUSH logs of experiments run on GENI slices. In addition to extracting more detailed provenance from GUSH logs, version 2.5 of the NetKarma plug-in includes richer annotations of the provenance visualized and is more configurable by the user - allowing greater flexibility in visualizing and comprehending the provenance of an experiment. Version 2.5 also includes the NetKarma GMOC Adaptor which automates the annotation of NetKarma provenance graphs with network data from GMOC.

The additional capabilities added to NetKarma in version 2.5 include:
- Enhancements to the NetKarma Adaptor used to process GUSH logs and ingest the provenance of GENI experiments run on GUSH into the NetKarma Provenance System. This version handles a broader range of GUSH logs.

- Enhancements to the NetKarma Cytoscape plug-in for NetKarma that includes more detailed provenance and the addition of annotations to processes, artifacts, and the edges that connect the processes and artifacts in the OPM graphs generated using NetKarma. This version of the visualization plug-in also has increased configurability by the user as to the details displayed when visualizing provenance graphs.

- The new NetKarma GMOC Adaptor is being released which extracts additional metadata from the GMOC database and annotates the GUSH provenance graphs with data from the GMOC database. This Adaptor is configured to run as a background process and automatically annotates the provenance ingested into NetKarma without any additional steps required by GENI users.

The NetKarma Provenance System and the NetKarma GMOC Adaptor are running as a persistent service on servers at Indiana University. This allows GENI users to process a GUSH log through the NetKarma Adaptor and visualize the resulting provenance graph, including the automated GMOC annotations, without having to install the NetKarma server. To try out the NetKarma provenance visualization, GENI users can process their own GUSH logs or the sample log file included with the NetKarma adaptor.

Project Participants

During this time period, participants in the NetKarma project included: Beth Plale, PI of School of Informatics and Computing, Chris Small, Co-PI of InCENTRE, Scott Jensen, Postdoctoral Fellow, Peng Chen, PhD student, Yuan Luo, PhD student, Robert Ping, Project Manager in Data to Insight Center that Professor Plale directs.

Publications & Documents

NetKarma Adapter Version 2.5 User Manual
http://pti.iu.edu/sites/default/files/netkarma-gush-adaptor_v2.5_user_manual.pdf

NetKarma Provenance Retrieval and Visualization Plug-in Version 2.5 for Cytoscape User Manual
http://pti.iu.edu/sites/default/files/netkarma_provenance_retrieval_and_visualization_plug-in_for_cytoscape_user_manual_v1.1.0.pdf

NetKarma Version 2.5 Quick Start Guide
http://pti.iu.edu/sites/default/files/netkarma_version2.5_quick_start_guide.pdf

Plan for Evaluating NetKarma Provenance System
http://groups.geni.net/geni/attachment/wiki/netKarma/GENI%20-%20Provenance%20System%20Evaluation.pdf

Provenance Metadata and Extensibility of Metadata Describing Measurement Data
http://groups.geni.net/geni/attachment/wiki/netKarma/GENI%20-%20Provenance%20System%20Evaluation.pdf

Collaborations and Planned Activities

At GEC12, the NetKarma Provenance team continued its involvement with the Instrumentation and Measurement team led by Harry Mussman, with Scott Jensen and Chris Small participating in the I&M working group session on Wednesday morning. At the I&M meeting, we presented slides as part of the "lightening talks" on current projects and our presentation focused mainly on our thoughts regarding Metadata as part of on-going MDOD development within the I&M group. The slides are available from the I&M group's GENI wiki: http://groups.geni.net/geni/attachment/wiki/GEC12InstMeasWorkingSession/GEC12-Provenance.pptx

The I&M group also formed teams for priority topics, with Chris Small participating in the "I&M Use Cases for Infrastructure Measurement, and Support for Operators" group, the "GENI User Tools and Services" group, the "Digital Object Archive (DOA) Service" group and both Chris Small and Scott Jensen participating in the "Descriptors, Objects and Registries, and Lookup Service" group.

We also discussed the MDOD and metadata with Deniz Gurkan of the University of Houston and representatives of Infoblox who had a demo on using IF-MAP to capture metadata. We also discussed the MDOD with Giridhar Manepalli of CNRI who is heading the "Descriptors, Objects and Registries, and Lookup Service" I&M team. Following Mark Berman's outbrief on the experimenter's working group meeting, we discussed with him what information experimenters were looking for regarding the environment their experiment ran in. There had been a brief discussion of the topic in the experimenter's working group and Mark was able to share some high-level thoughts on the topic.

We also talked with Jeffery Hunker, following up on discussions we had at GEC11 during his session on attribution. There are overlaps and possible synergies between attribution in GENI and provenance of experiments as captured by NetKarma.