NetKarma Quarterly Report 01/01-03/31/2010

Beth Plale – Principal Investigator

Chris Small – Co-Principal Investigator

Summary

- *Demonstration of NetKarma*
- *Gush provenance capture*
- *Karma 3.0 released*

Major Accomplishments

Milestones Achieved

<u>Demonstration of NetKarma:</u> Demonstrated a working version of NetKarma ingesting provenance information from GUSH. We visualized the provenance graph that is generated when an experimenter uses Gush to manipulate their experiment. Delivered at GEC-7

NetKarma Poster used at the demo session can be found here:
http://groups.geni.net/geni/attachment/wiki/netKarma/Netkarma_poster_gec7.pdf

Description of Work Performed During Last Quarter, Q2

Activities and Findings

<u>GUSH provenance capture code released</u>: We have released version 1.0 of the "gush2netkarma" software. This code contains tools to parse log files generated by Gush. While this software allows injestion of Gush logfiles into the Karma provenance framework, it also provides a generalized rules framework that can be used to generate provenance information in other GENI projects. We hope to use this framework in capturing the status and workflow of GENI experiments. The framework developed should allow us to capture artifacts and instrument additional GENI components easily.

The gush2netkarma-1.0 code is linked from the NetKarma GENI WIKI page and available at the following location:
https://globalnoc.iu.edu/grnoc-internal/file-bin/system-engineering/software-projects/netkarma/gush2netkarma-1.0.tar.zip

Karma 3.0 Released: The new release of the core Karma system contains instrumentation using Axis2 handlers, more extensive test clients, and better documentation. Karma v3.0 supports provenance activities published from services, workflows and nested workflows. The provenance data is efficiently stored in a relational database, and supports Open Provenance Model (OPM) v1.0 standard for interfacing with the tool.

Project Participants

During this time, key participants in the NetKarma project included:

Beth Plale, PI
Chris Small, Co-PI
Girish Subramanian, PhD student
Yiming Sun, PhD student
Devarshi Goshal, PhD student
Sharanya Chinnusamy, MS student
David Ripley, technical staff
Robert Ping, Project and Information Management

Publications & Documents

NetKarma Status Update during GEC7 Planetlab cluster meeting
http://groups.geni.net/geni/attachment/wiki/netKarma/NetKarma_PL_cluster_GEC7_slides.pdf?format=raw

NetKarma Poster used at the GEC7 demo session:
http://groups.geni.net/geni/attachment/wiki/netKarma/Netkarma_poster_gec7.pdf

gush2netkarma-1.0 release:
https://globalnoc.iu.edu/grnoc-internal/file-bin/system-engineering/software-projects/netkarma/gush2netkarma-1.0.tar.zip

Collaborations

We have worked extensively with Jeannie Albrecht, lead of the Gush team, to ingest provenance relevant information from log files produced by Gush into the Netkarma system. Professor Albrecht has been extremely cooperative in providing assistance in understanding the various data artifacts produced by Gush and in deploying a demonstration instance for the development of the "gush2netkarma" software.

We continue to work with the Gush team on refining the provenance data collection process. This includes adding hooks to the Gush code to facilitate the collection of artifacts from the Gush processes.

We have had discussions with John Hartman and Scott Baker of the Raven project to discuss the workings and workflow of the Raven provisioning service. We are studying the logs and information presented through the Owl monitoring service for collection into NetKarma.

We have started initial discussion with the GMOC. We are studying the extent to which operational data collected by the GMOC, especially control plane related data, could be collected from the GMOC repository instead of the original source. The GMOC repository would provide a unified, normalized schema to query against compared to widely differing data and formats used by control plane and other data sources. Collecting data already obtained from various sources and stored in the GMOC database schema may speed and simplify the work need for data collection.

We had a preliminary discussion of how OpenFlow state such as flow tables and controller policy rules can be used. We focused on collection from the Indiana University OpenFlow campus trial infrastructure as an initial test but techniques should be applicable to all OpenFlow deployments.

Planned Activities for Q3

We will work to establish connections with relevant projects in the Control Plane and Measurement Plane areas. We plan to make decisions of what Control plane information we will integrate into the Netkarma data collection during Q3.

We will continue to develop software to allow the integration of Raven metadata into Netkarma. This may include data such the locations, success of the distribution and time of deployment of software package to all nodes in an experiment. References to the software itself on the Raven repository would also be incorporated in to the Netkarma description of an experiment. The software reference could be very useful if future creation of the ability to recreate an experiment from information captured in the Netkarma provenance trace.

.