NetKarma
Quarterly Report for Period Ending GEC11

Beth Plale, School of Informatics and Computing, Indiana University
Chris Small, Global Research Network Operations (GMOC), Indiana University

Summary

For the quarter ending with GEC11, we augmented the existing provenance captured by NetKarma with network topology information from the GMOC Global Research NOC database and network measurement data from PerfSONAR-PS PingER service. These were identified in a GEC10 milestone as an additional source of provenance. Additionally we continue to refine visualization of provenance to make it more useful for experimenters. We built a plug-in for CytoScape and in this quarter enhanced it to display the additional provenance annotations being captured and enable greater configurability of details to display. These developments were highlighted in a poster presented at GEC11.

Through discussion with Harry Mussman at GEC11, we identified two key engagement points of our group with the instrumentation and measurement team: 1.) our team will assist in the definition of the transaction record MDOD schema; 2.) we will also explore the representation of provenance as an object in the instrumentation repository. Through discussion with Vic Thomas and Mark Berman on the benefit of using one of the current GENI experiments from which to capture provenance, we agreed to explore a WiMAX DoS application being carried out by Lu Yu under direction of professor Richard Brooks, both at Clemson.

Quarter Deliverables Discussed

S3.d Updated plan for making provenance information available to experimenters: The updated plan for users to submit their GUSH logs and get back a visualizable provenance graph documents details of the submission process; including the URL of the persistent Axis2 web service where donators of GUSH logs can programmatically download their provenance graph and then visualize it using the CytoScape visualization tool.

S3.e GEC11 demonstration and outreach: At GEC11 we presented "Experiment Provenance: Towards Links to Network Measurement Data" poster that showed annotations to a NetKarma provenance graph using network topology from the GMOC database and measurement data from PerfSONAR PingER service. Due to the change in venue for GEC11, the demo session was cancelled, so a poster session was the only venue available to us to satisfy the deliverable.

S3.f Deliver software and documentation (Due 08/26/11): The upcoming end of August release of NetKarma will include 1.) the Karma provenance system, 2.) visualization plugin to cytoscape, 3.) adaptor for extracting provenance from GUSH, 4.) adaptor for extracting network topology data from the GMOC database, and 5.) adaptor for capturing PingER data and adding it to netKarma.

Description of Work Performed Prior Quarter

At GEC10 we identified the GMOC database as a source of additional metadata on the network topology used to run GENI experiments on PlanetLab slices. During this quarter, the NetKarma plug-in to the CytoScape visualization tool was enhanced to display these metadata as a pop-up dialog when a user clicks on a process node in the provenance graph. Additionally, instrument and measurement data was identified as another potential source of metadata annotations. As a prototype of capturing measurement data, we configured GUSH experiments that performed a distributed experiment on a PlanetLab slice to capture network measurement data using the PerfSONAR-PS PingER module and then harvested that metadata using a NetKarma adaptor to annotate the edges between processes in the provenance graph. As with the GMOC data, the NetKarma plug-in for CytoScape was enhanced to query the Karma database when a user clicks on an edge in the provenance graph and display the measurement data annotations in a pop-up dialog.

The additional annotations added to the provenance graph are captured by a harvesting module added to NetKarma that runs periodically during a GUSH experiment and determines which new edges have been added to the provenance graph since the last iteration. The NetKarma adaptors for the GMOC database and measurement archive are then triggered to retrieve metadata and submit provenance notifications to Karma which are then added to the OPM graph generated by NetKarma.

The NetKarma adaptor that processes the GUSH logs to generate Karma provenance notifications was enhanced to capture provenance between child processes in GUSH experiments. The NetKarma software was also enhanced to enable metadata annotations to be added for a wider array of data (artifacts in OPM) so that not only files could be annotated, but also "data blocks" which capture data written to standard output by the programs executed in a GUSH experiment. The Karma software was also extended to allow annotations on the edges between processes. The NetKarma plug-in for CytoScape was enhanced to provide greater configurability by scientists as to the annotations and details that would be displayed for each node in the provenance graph displayed in CytoScape.

At GEC11 we presented a poster that illustrated the enhancements made to NetKarma during this quarter to harvest additional metadata regarding the network topology of an experiment's nodes from the GMOC database and periodically retrieve measurement data during the running of an experiment to annotate the provenance graph with data captured using PingER. The poster also presented screen captures of the enhanced Karma CytoScape plug-in that illustrated the new annotations of the provenance graph nodes and edges.

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 GMOC, Mehmet Aktas, Postdoctoral Fellow, Scott Jensen, Postdoctoral Fellow, Devarshi Goshal, PhD student, Peng Chen, PhD student, You-Wei Cheah, PhD student, Yuan Luo, PhD student, Robert Ping, Project Manager in Data to Insight Center that Professor Plale directs.

Mehmet Aktas is beginning a faculty position in Turkey August 2011. His commitment to the project will be missed. Scott Jensen, PhD is taking his place.

Publications & Documents

Karma Adaptor Tool V 2.0 User Manual.

GENI Provenance Registry. Updated plan for making provenance information available to experimenters.

Experiment Provenance: Towards Links to Network Measurement Data, Poster presented at the 11th GENI Engineering Conference.

Unmanaged Workflows: Their Provenance and Use, Mehmet Aktas, Beth Plale, David Leake, Nirmal K. Mukhi, book chapter under review, July 2011

Key Provenance of Earth Science Observational Data Products, Mehmet Aktas, Beth Plale, Helen Conover, Prajakta Purohit, Technical Report, Data to Insight Center, Indiana University, July 2011

Collaborations and Planned Activities

At GEC11 we participated in discussions with the Instrumentation & Measurement team on their development of a Measurement Data Object Descriptor (MDOD) to contain the metadata describing measurement data. Following up on these discussions, we are collaborating with them on the prototyping of the section of the MDOD that captures provenance and distribution of the measurement data. In the I&M working session at GEC11 we presented a "lightning talk" on NetKarma and goals for annotating provenance with additional measurement data.

At the I&M working session we discussed the capture of provenance in distributed systems and NetKarma's approach to provenance capture with Max Ott of NICTA in that identifying the completeness of provenance is an issue in his research.

Following discussions at GEC10, in the GEC11 Workshop on Attribution in GENI, NetKarma was mentioned as one of a handful of projects in GENI that address attribution. Following the workshop we also talked with Jeffrey Hunker about provenance, attribution, and their efforts to create an ontology of attribution.

At GEC11 we discussed possible projects in GENI that would benefit the most from provenance capture with Mark Berman of the GPO and following up on those initial discussions we initiated discussions with Richard Brooks regarding the work his team is doing on researching DDoS attacks using WiMAX. As another avenue for making the GENI community aware of NetKarma, Heidi Dempsey of the GPO suggested the possibility of incorporating it into a GUSH tutorial at a future GEC conference.