Plan for Evaluating netKarma Provence System

In GENI, the provenance captured has three dimensions that should be evaluated. Drawing from previous information quality literature [1], these dimensions are accuracy, completeness, and relevancy, and are defined for data provenance and explained in more detail below.

1. Accuracy - the extent to which provenance capture and representation are correct and free-of-error.
2. Completeness - the extent to which provenance is not missing and are of sufficient breadth and depth for the task at hand.
3. Relevancy - the extent to which the provenance is applicable and helpful for the task at hand.

Selected application: We have selected two applications for the evaluation of the netKarma provenance system. These two applications are representative of the breadth of experiments that may be executed on GENI. The first application is an experiment of Clemson University's WiMAX Research Project (http://groups.geni.net/geni/wiki/FirstGenBrooks), which focuses on assessing distributed denial of service vulnerabilities on WiMAX [2]. WiMAX [3] is an IEEE 802.16-2004 standard for fixed wireless applications and 802.16e-2005 for mobile wireless. Using the source code provided by the Clemson team, we have the experiment running and are capturing provenance regarding the network topology, configuration parameters, and network traffic information. The provenance captured is being ingested into NetKarma and we are using Cytoscape and the NetKarma plug-ins we have developed for Cytoscape to visualize the provenance for these experiments. We have initiated discussions with the Clemson team on how this provenance would be most beneficial to GENI experimenters.

The second GENI application that we are working with is an experiment from the Phoebus project [5] that uses the eXtensible Session Protocol (XSP) [4] (http://groups.geni.net/geni/wiki/FirstGenSwany). The XSP experiment focuses on experimenting with different network configurations and evaluating how it influences network performance. In this experiment, we capture provenance such as network bandwidth and latency. Provenance for both experiments is captured through adaptors designed for their respective network environments. We have had discussions with members of the experiment team and are using GENI resources in multiple locations to run the experiment and determine the provenance that can be captured.

Provenance model: The provenance model for our selected applications is as follows. Clients/services are modeled as processes, using terminology from the Open Provenance Model (OPM). Parameters of these clients/services are modeled as annotations of their respective processes. Packets that are sent between services are represented as artifacts and their respective sizes are stored as attributes of these artifacts. The edges in the provenance model represent the relationship amongst the clients/services. Parameters like bandwidth that are related to the transfer processes are stored as annotations of these edges.

Assessment Criteria:

Criteria	What	How manifested
Relevancy	Relevancy to whom: experimenter (1. Clemson researchers, 2. XSP experimenters)	Difference between two executions: attributes and trace structure
Relevancy	Relevancy to whom: researchers who reuse a network trace	Trace level completeness
Accuracy	Duplicates do not conflict; range checks on parameters. Ok for this to be syntactic checks instead of both syntactic and semantic.	Data is accurate and has no conflicts
Entity level completeness	Whether different parameters are available for each node in a provenance trace.	Complete data for a node in the environment is captured
Trace level completeness	Completeness of an experiment trace. That is, a trace has all the nodes and edges of an experiment.	Nodes and edges for an experiment are complete and represented correctly

References

[1] Y. W. Lee, D. M. Strong, Knowing-why about data processes and data quality. *Journal of Management Information Systems, 20*(3), (2003-2004), 13-39.

[2] J. Deng, R. Brooks, J. Martin, "Assessing the Effect of WiMAX System Parameter Settings on MAC-level Local DoS Vulnerability", to appear in International Journal of Performability Engineering, (2011).

[3] WiMAX, http://www.wimax.com (last accessed Jan, 2012)

[4] E. Kissel and M. Swany, The eXtensible Session Protocol: A Protocol for Future Internet Architectures, *in review*

[5] Phoebus project, http://damsl.cis.udel.edu/projects/phoebus/ (last accessed Jan, 2012)