# Visual Exploration and Analysis on Hosts, Users and Applications in Enterprise Networks

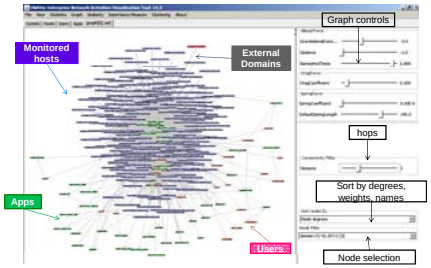
**Qi Liao, Dirk VanBruggen, Andrew Blaich and Aaron Striegel.**
**Department of Computer Science and Engineering, University of Notre Dame, U.S.A.**

## Problem

This work is motivated by a relatively innocuous question: *how well do we know our network*? Although there exists a wide spectrum of security analysis and network visualization toolkits, the tools tend to focus on *where* and how much communication is occurring, not *whom*, *what* and *why* are the communications occurring. With the increasing trend towards distributed systems and the ever changing behavior of users, the *context* of the communications with regards to security and enterprise management rather than packet/flow *content* becomes more important than ever. As a result, troubleshooting and security analysis devolve into the equivalent of a digital spelunking expedition, frequently overwhelming the network administrator and resulting in considerable lost enterprise productivity or increased security risk.

## Solution

We argue that for enterprise networks, knowing end-to-end connections is too coarse to be useful and *network addresses* and *port numbers* become less useful identifiers for visualizing network activities. We believe that the inclusion of relatively simple *context* (*users*, *applications*, and *data*) in addition to host locations coupled with advanced data analysis techniques can shed significant light on the question of *what is really going on in my network*?

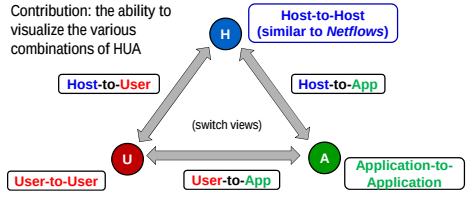



To that end, we created *ENAVis* (*Enterprise Network Activities Visualization*), a graphical tool that brings the notion of local context (*whom*, *what*, *why*) to dramatically improve how administrators view the network. The key innovation of *ENAVis* is to leverage local context to allow the administrator to quickly assess relationships among the hosts, users, and applications using the network. The powerful, yet intuitive interface of *ENAVis* enables administrators to seamlessly browse, assess, debug, and analyze the timelines of activities within the network on the order of seconds, whereas existing tools require hours if such tasks are even possible.

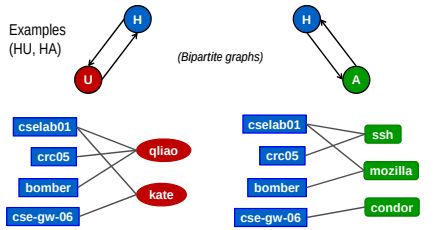
Q. Liao, A. Blaich, A. Striegel, and D. Thain, "*ENAVis: Enterprise network activities visualization*." in Proceedings of the USENIX 22nd Large Installation System Administration Conference (LISA '08), San Diego, CA, November 9-14 2008, p. 5974.

## Hosts-Users-Applications (HUA) Graph Model

Contribution: the ability to visualize the various combinations of HUA

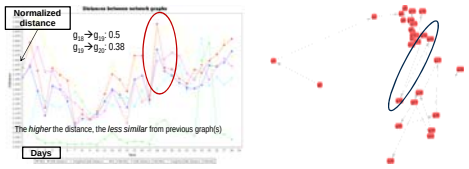



Examples (HU, HA)



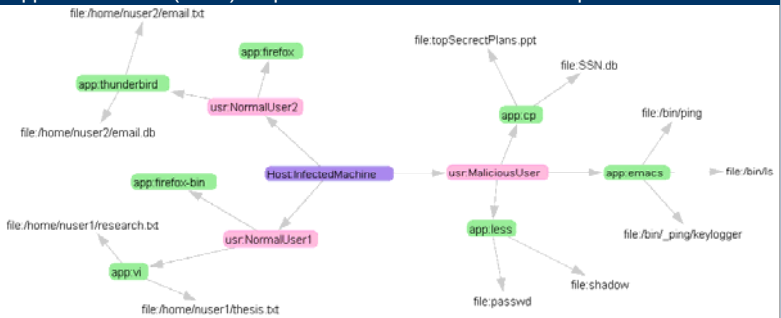

## Visualize Similarity of Network Graphs

- *ENAVis* helps human investigator to answer the following questions: How *similar* (or equivalently how *different*) from day-to-day network activities? What are the *variance* and *invariance*? What changes are *normal* / *abnormal*? How to *visualize evolution* of changes?
- *Similarity visualization* includes graph sizes, diameter, degree distribution, max/min common sub/supgraphs, inter/intra graph clustering, graph edit distance, multi-dimensional scaling, etc.
- *ENAVis* links effective data mining techniques to enterprise networks visualization, and brings the gap between daily network monitoring and high level decision making.



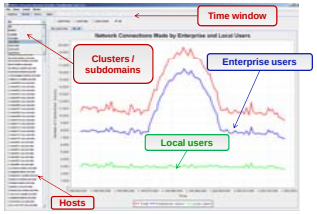

## Host-Users-Applications-Files (HUAF) Graph Visualization and Interactive Exploration

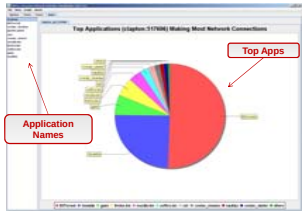
The *local-context* associated with each network connection can include *any* relevant information, including but not limited to *users*, *applications*, as well as *data* files they touched.

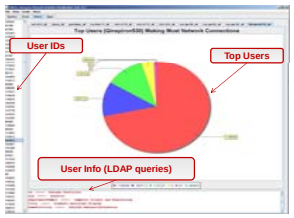



## Network Activities by Top Responsible Users and Applications

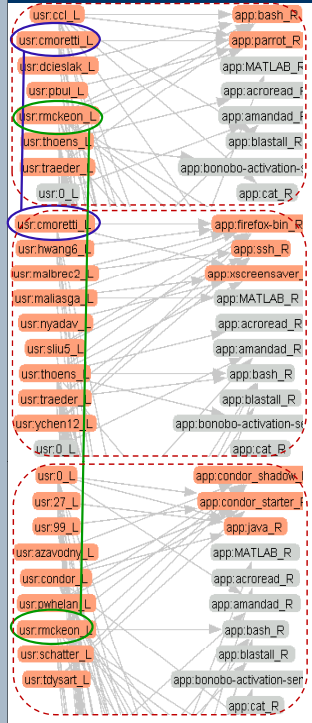
- Network activities in a user specified time window, within a subnet of hosts.
- Admin can drill down the network connections by *the type of users* (i.e., enterprise vs. local).
- Pie charts are automatically generated for the top *N* most active users making most of network connections and the top *N* applications run by the user in question.
- Being able to drill down further to *who* is responsible in terms of users and applications (not just hosts) can be useful in case of forensics and policy compliance auditing.
- Network operators can **quickly pin down the problem source** with just a few mouse clicks with a minimal effort and time.




## Bipartite Community (HH, HU, UA, AH)




**Community detection** or **clustering**: proved be very useful in understanding the networks, and have applications in traffic classification and classic anomaly detection.

Clustering heterogeneous ENAVis graphs by separating into *bipartite graphs*.

Example: *User-App* grouping containing three communities.

One interesting observation is the phenomenon **overlapping communities** with those important users bridging between two or more clusters.

## Conclusion

- It is important to include most dynamic components, i.e., hosts, users, applications and data in network monitoring, visualization, and analysis.
- ENAVis collects, correlates, visualizes, interactively explores, and analyzes the above missing context information associated with each network connection.
- Novel application of multidimensional views, visualization techniques, data mining and machine learning algorithms, and graph theory can significantly improve administrators' understanding and insight on their networks.

## Contact information

**Website**: Full demonstration movie of tool walkthrough and viewer download available at http://netscale.cse.nd.edu/Lockdown
Department of Computer Science and Engineering
384 Fitzpatrick Hall, University of Notre Dame,
Notre Dame, Indiana 46556
Tel: 574.631.6896, Fax: 574.631.9260, Email: qliao@nd.edu