Network Security and Anomaly Visualization

Qi Liao, Department of Computer Science, College of Science and Technology

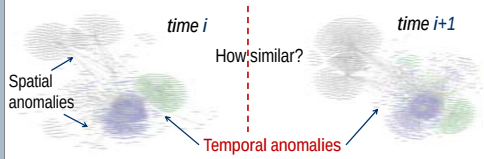
CMU CENTRAL MICHIGAN UNIVERSITY

(Science, Engineering and Technology Day, 2011)

Problem

Large-scale enterprise networks typically involve thousands of users and applications, terabytes of data, millions of connections. How do we know precisely what is going on in our network? If there are intrusions and attacks, can we quickly detect and find who is responsible? Some attacks are noisy and easier to detect, such as port scans and DDoS. What about less obvious intrusions from advanced and persistent attackers? A few key questions to ask:

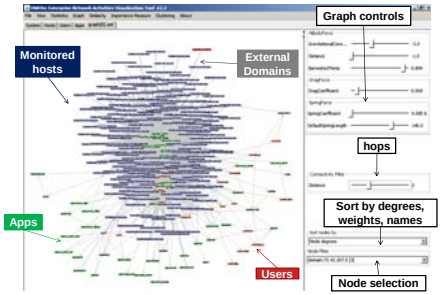
- What are the *changes* happened in networks?
- What are the *variance* and *invariance*?
- How *similar* (or *different*) from day-to-day network activities?
- What changes are *normal* / *abnormal*?
- How to quantify and visualize the *evolution* of changes?



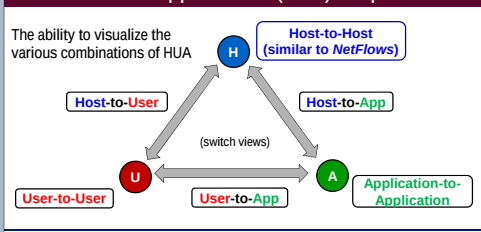
Solution

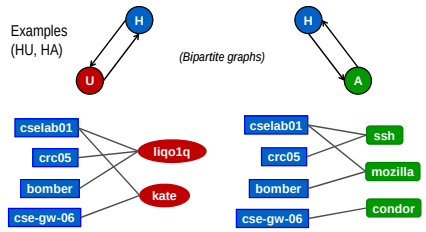
First, we combine the traditional logging (Netflow, firewalls, tcpdump, IDS, syslog, etc), which gives a higher level of connectivity (IP addresses and port numbers), with finer granular logging (users, applications, data, etc). We correlate the events together to form *local context* of network connections.

Second, we develop a user friendly, highly interactively visualization tool (formerly know as ENAVis) that allow network operators and administrators to quickly explore the vast amount of log data, which provides both situation awareness and details on demand (overview + context). Most importantly, the tool combine the advantages from both automatic graph data mining algorithms and manual visual analysis from expert domain knowledge.



Hosts-Users-Applications (HUA) Graph Model





Graph Distances

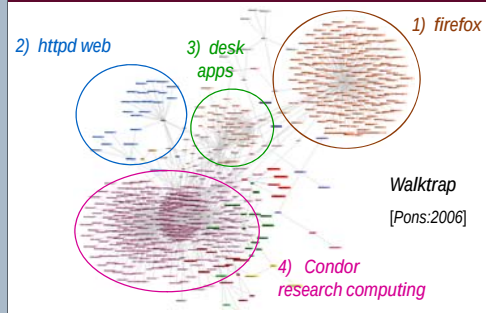
- *Graph Edit Distance* (GED) [Bunke07] to measure the graphs' similarities.
- Number of operations required to transform from one graph to the other.
- Maximum common subgraphs (MCS) based:

$$d(g_1,g_2)=1-\frac{|mcs(g_1,g_2)|}{\max(|g_1|,|g_2|)}$$

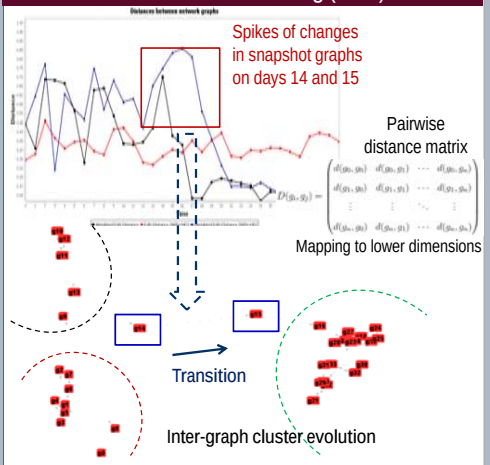
- Graph edit distance (GED) based:

$$d(g_1,g_2)=\frac{|g_1|+|g_2|-2|mcs(g_1,g_2)|}{|g_1|+|g_2|}$$

Intra-graph Community Visualization

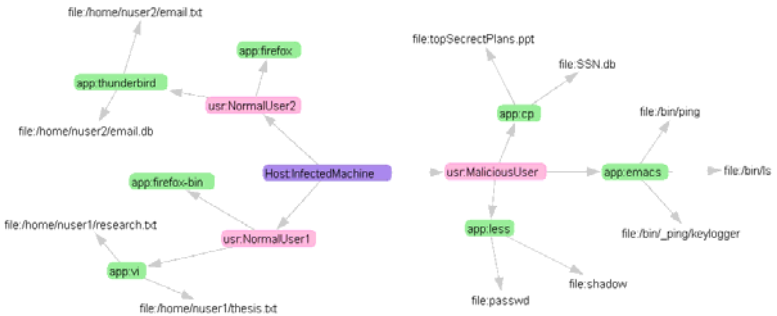


Multi-dimension Scaling (MDS)

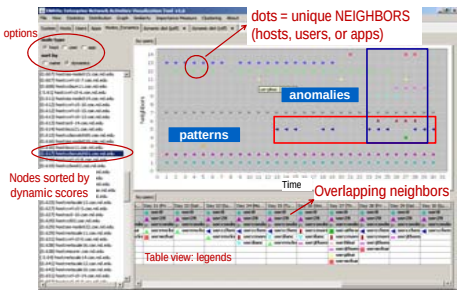


Host-Users-Applications-Files (HUAF) Graph Visualization and Interactive Exploration

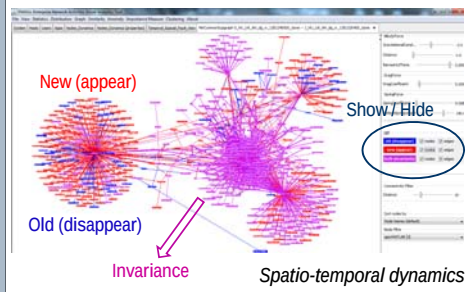
The *local-context info* associated with each network connections can be expanded to include *any* relevant information that is appropriate to a specific organization.



Node Similarity Visualization



Graph Differential Visualization



Publications

- Qi Liao, Andrew Blaich, Dirk VanBruggen, and Aaron Striegel. Managing networks through context: graph visualization and exploration. *Computer Networks, Special Issue: Managing Emerging Computing Environments*, 54(16):2809-2824, November 15 2010.
- Qi Liao, Aaron Striegel, and Nitesh Chawla. Visualizing graph dynamics and similarity for enterprise network security and management. In *ACM Proceeding of the 7th International Symposium on Visualization for Cyber Security (VizSec'10)*, pages 34-46, Ottawa, Canada, September 14 2010.
- Qi Liao, Andrew Blaich, Aaron Striegel, and Douglas 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. USENIX BEST PAPER AWARD.

Contact information

Dr. Qi Liao
Pearce Hall 417
Department of Computer Science
Central Michigan University
Mount Pleasant, MI 48859
Tel: 989.774.4419, Fax: 989.774.3728
Email: qi.liao@cmich.edu

We want smart students like you to work our research projects. Visit http://cps.cmich.edu/liao1q/research for more information.