3D Anomaly Bar Visualization for Large-scale Network

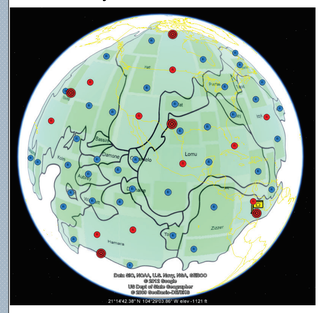
IEEE Conference on Visual Analytics Science and Technology (VAST) Challenge Workshop, Seattle, Washington, October 14, 2012

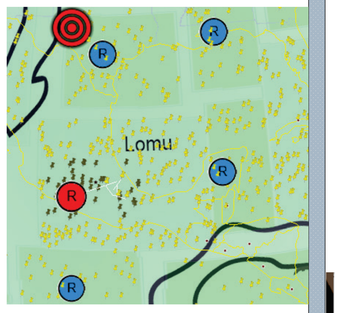
Tao Zhang (Central Michigan University), **Qi Liao** (Central Michigan University) and **Lei Shi** (Chinese Academy of Sciences)

Problem

The VAST 2012 mini-challenge 1 data embodies a large-scale enterprise network with both network traffic log and geographic information. The log data coming from locations all over the Bank of Money facilities that contains close one million IP addresses. The main goal for our visualization solution is to create large-scale cyber situation awareness.

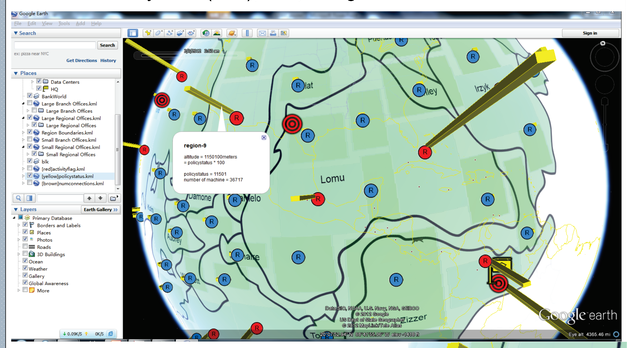
From this point of view, two problems become the main impediments to detect operational changes outside of the norm. First, geographic factors [1] may help to detect anomalies and investigate; and second the time dynamics could also be important for analysis. Therefore, how to represent anomalous network behavior with the geo-information is the critical part of the task. In addition, big volume of the data could be another challenge for human beings visual analysis. There are nearly 900,000 IP addresses and over 4000 physical locations. What and how to represent dominates the efficiency of analyzing anomaly in such an enormous network.

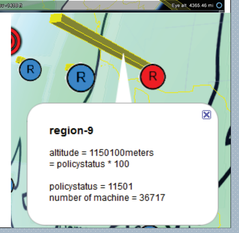




VC12 MC1 Data Description by Google Earth

We propose a useful method to simplify the visual representations by data aggregation and represent network status data by 3D bars together with position information. Our visualization solution is based on an existing geographic information system (GIS), i.e., Google Earth.

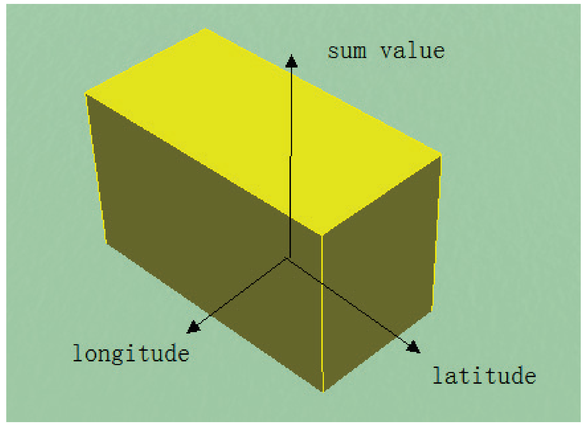




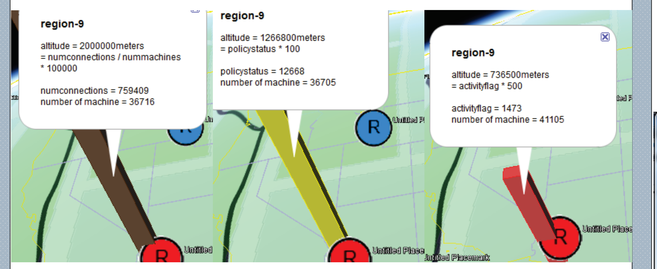
The UI: the left panel includes various filters for the result files based on different calculations; the right window is 3D visualization with interactions such as bubble with detail information by clicking.

3D Anomaly Bar Visualization

We choose the 3D bar representation because the extra height dimension makes the spatial data distribution more intuitive than a flat 2D color-coding. Besides, we use distinct colors of the bars to distinguish different attributes in concern. Investigators could view single or multi scenarios by clicking the file filter in Google Earth. Animated visualization is used to connecting the dots between timelines.

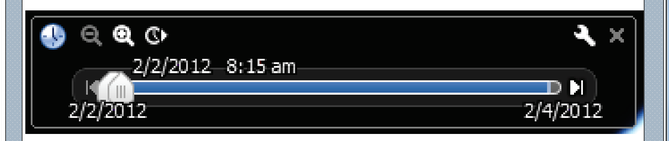


We choose each IP address' latitude and longitude as the bar's position, and define each bar's altitude by the sum value it contains.

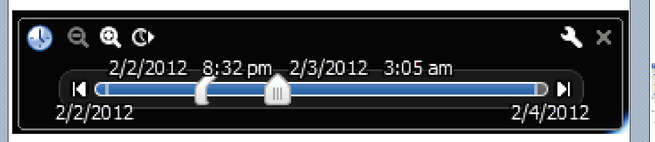


# of connections **policy** **activity**

Using *colors* and *heights* to distinguish different bars with their different concerned attributes.



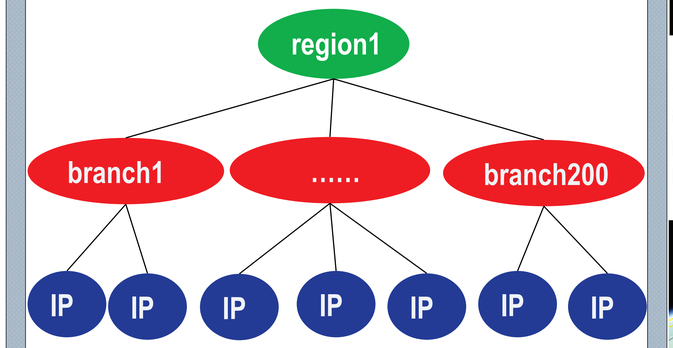
With adding timestamp to each item. An modifiable animation is available.



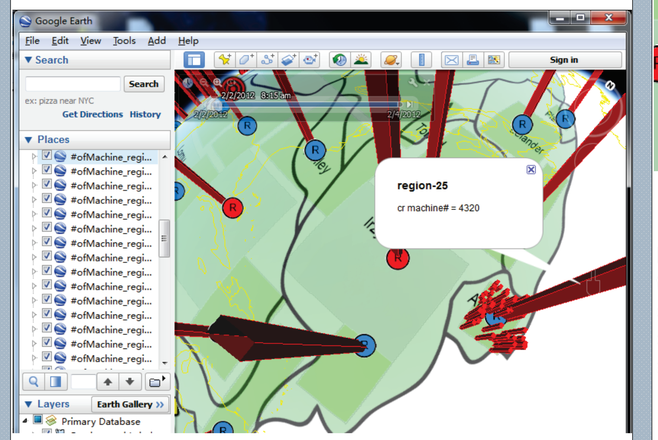
Animation can be also viewed by setting a particular time zone with a start and end time

Data Processing

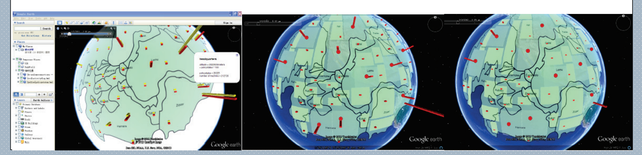
We mainly use *regions* as the aggregation granularity. While rendering at finer granularity is useful when focusing on a problem source, using branch, for example, will generate more than 4000 3D bars for whole BoM world, which will hardly be displayed and recognized by human beings. However, grouping by branch will be an appropriate way to represent more details when studying only a few regions. By using BMT time as another dimension, we obtain a time series of status distributions, one per 15 minutes. We firstly show one bar per region, corresponding to either the connection, activity, policy sum or number of machines. We may also calculate more in detail after detecting any region's status anomaly. Selecting KML files together may make a visualization combined by both regions and branches.



Regions are also divided by size (big / small). Different IP addresses can be different types of machine (ATM, teller machine etc.)

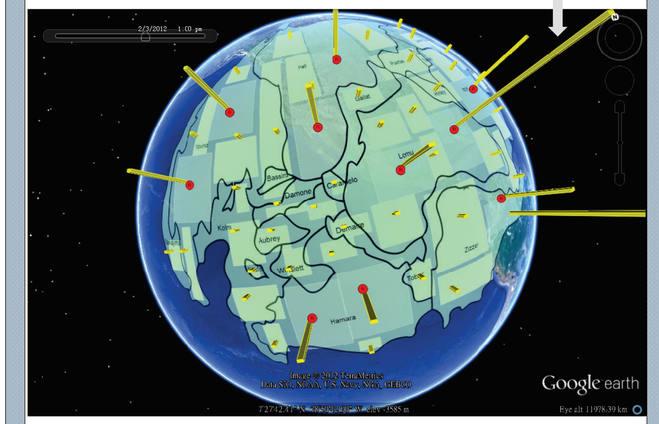


Generally speaking, the higher the bar and/or the larger the difference of bars, the more anomalous of the network is.

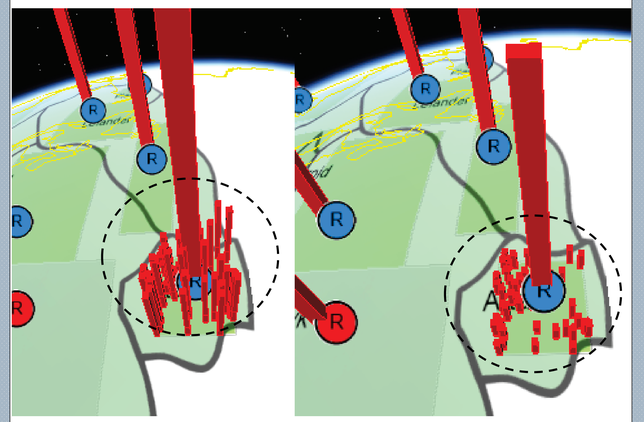


Preliminary Results

High bars indicate big policy violations (anomalies)



The height of policy status bars represents how severe a policy violation of a network (region, branch, etc) is. The rising patterns of the policy status can be a critical issue for BoM. It seems that most policy deviation warnings still exist over time, while the sum of policy status keeps increasing.



# of IPs online before 10am **# of IPs online after 10am**

Machines in Region 25 went offline starting at 10am and continued throughout the evening. Number of machines along the southeast coast of Bank of World had a significant reduction that is determined to be inside sets of bounding geo-coordinates are affected.

A video showing interaction for other answer can be viewed from http://cps.cmich.edu/liao1q/video/CMICH-Zhang-MC1.wmv

References

[1] Basak Alper, Selçuk Sümengen and Selim Balcisoy Dynamic visualization of geographic networks using surface deformations with constraints. In Proceedings of the 25th Computer Graphics International Conference (CGI'07), Petropolis, Brazil, May 30-June 2, 2007.