Status Update

Since submitting the proposal, I've finished the following tasks:

1. Set up a MySQL database and inserted the datasets from the VAST challenge. I've then processed the data into a form needed by my proposed solution. Because the original data has an entry for individual machines and individual status reports, I've aggregated them down to a much smaller dataset in terms of facilities and regions.
2. I've set up a basic node.js server and connected it to the database.
3. I've created the html for the website's page with all of the required static elements, including the map for the static view.
4. I've finished implementing one of the static view options, in which one of the attributes of the machines are encoded by colour.
5. I'm currently in the process of implementing one of the dynamic view options, in which one of the attributes of the regions are graphed as a function of time below the static view.

Not too many obstacles have been encountered so far. A small hitch I've come across is that due to the need to resize the map image provided by the VAST challenge, the resulting image is becomes very pixelated and ugly. I am currently considering getting rid of the background map altogether and simply drawing the regions against a blank background. Several of the other solutions from the VAST challenge have used this method and seem to provide a much cleaner image. Other than this, little changes have been made to the original plan.

Related Work

In this section, past applications for visualizing computer network health will be discussed. In particular, I will focus on the applications which were submitted as part of the 2012 VAST mini challenge. Looking at three key design decisions, I present the decisions made by these other applications and compare them against the decisions made by the current system.

How do we visualize changes in a network over time?

In visualizing changes in a network over time, past solutions have generally taken one of two approaches: (i) provide a timeline showing the value of a network attribute as a function of time or (ii) provide an animation which shows a frame by frame evolution of the network over time. Solutions which have gone the first route [1-3], have the advantage of showing the evolution of multiple network attributes at once, allowing the user to discover correlations between attributes. In contrast, solutions going the second route such as that submitted by the BusinessForensics team [5], allows the user to view how individual regions evolve over time, rather than having their data be aggregated into network-wide summaries and consequently lose a lot of detail. At the same time, these solutions suffer from their inability to provide a clear overview of the network's change over time. Some solutions, such as that submitted by the Secure Decisions team [1], counterbalance the loss of information due to aggregation by allowing users to select a particular point in a network-wide timeline in which to drill down and break the aggregated attribute value down to its component parts (e.g. the number of machine connections per business unit). In this way, by providing the user with an overview of how the network changes over time, then allowing them to drill down into a particular timepoint for detailed information, the disadvantages of the two forementioned routes are mitigated. In the current system, I borrow from the Secure Decisions solution and provide an overview of the network's change over time in which users can select timepoints to drill down into. Unlike the Secure Decisions solution, however, I provide a breakdown of an attribute into regional contributions. In this way, the system is able to support a back and forth "drilling down" between the static and dynamic views, whereby a user selects a particular timepoint to show in greater static detail, then using this greater level of detail users select a particular region for which to show isolated dynamic information.

How do we show the various network attributes?

In providing snapshot of the network's attributes at a given point in time, most solutions have tended to encode the values of machine attributes onto a geographic map of the network so as to match the user's mental model of the network as it exists in the world. The variability in solutions have mainly come in the details of how these attributes are encoded on the map. Several solutions provide a panel of attributes which they can turn on or off [2] which then maps the various selected attributes simultaneously onto the static view using separate channels of encoding. Others have instead reserved the static view solely for a single attribute, such as changes in policy levels, and show other attributes by other, time-dependent views [1]. The current solution has favoured the former approach, although with the difference that rather than providing a panel of checkboxes to view multiple attributes simultaneously, it instead provides a set of radio buttons used to toggle between the various attributes. The obvious disadvantage with this appraoch is that finding correlations between different attributes becomes difficult, requiring users to constantly toggle back and forth between the attributes of interest. However, because the current system is less interested in finding relationships between attributes, and more interested in finding points of anomaly, the radio buttons were chosen for their added simplicity. Because, only a single attribute is ever shown on the page, the system is able to encode them uniformly, not requiring its users to learn the semantics of multiple encodings.

References

[1] M. Farry, R. Stark, A. Wollocko, and M. Borys, “CRA-Farry-MC1,” *2012 VAST Challenge*, 2012, http://hcil2.cs.umd.edu/newvarepository/VAST%20Challenge%202012/challenges/MC1%20-%20Bank%20of%20Money%20Enterprise%20Cyber/entries/Charles%20River%20Analytics/

[2] C. Horn, C. Ellsworth, and D. Halperin, “SecureDecisions-Horn-MC1,” *2012 VAST Challenge*, 2012, http://hcil2.cs.umd.edu/newvarepository/VAST%20Challenge%202012/challenges/MC1%20-%20Bank%20of%20Money%20Enterprise%20Cyber/entries/Secure%20Decisions/

[3] L. Laberge et al., “GDC4s-Laberge-MC1,” *2012 VAST Challenge*, 2012, http://hcil2.cs.umd.edu/newvarepository/VAST%20Challenge%202012/challenges/MC1%20-%20Bank%20of%20Money%20Enterprise%20Cyber/entries/General%20Dynamics%20C4%20Systems/

[4] R. Pabst, “BF-Pabst-MC1,” *2012 VAST Challenge*, 2012, http://hcil2.cs.umd.edu/newvarepository/VAST%20Challenge%202012/challenges/MC1%20-%20Bank%20of%20Money%20Enterprise%20Cyber/entries/BusinessForensics/

[5] J. Gbls-Szab et al., “SZTAKI-DMS: OWLAP Analytics Beta,” *2012 VAST Challenge*, 2012, http://hcil2.cs.umd.edu/newvarepository/VAST%20Challenge%202012/challenges/MC1%20-%20Bank%20of%20Money%20Enterprise%20Cyber/entries/MTA%20SZTAKI/