AGILE VISUAL ANALYTICS FOR BANKING CYBER “BIG DATA”

David Jonker, Scott Langevin, Peter Schretlen, Casey Canfield*
Oculus Info Inc.

ABSTRACT
This paper describes the rapid development of a tailored cyber situational awareness and analysis application for the 2012 IEEE VAST Mini-Challenge 1 (MC1) – Cyber Situation Awareness. The novel aspect of this project was in the process of developing the tailored solution for a “big data” application. Aperture is an open, adaptable, and extensible software framework, toolkit, human information interaction design team was able to determine what components of the solution were provided in the raw data, the most effective methods for aggregation and scaling of the data, and which components of the solution needed analytic services to derive answers from those aggregations of the raw data. For example, impact multipliers were designed that placed more importance on severe events that occurred higher in the bank organizational zone and organizational hierarchy. The next step involved sketching the design of the visualization layout, form factor and structure, incorporating a view of the entire bank enterprise, and accounting for differentiation of time zone and organizational hierarchy. Traditional visualization expressions were deliberately used to reflect Aperture’s target user group of developers without a visualization background. Finally, the color schemes for the policy violation expressions, the health expressions, and the derived “questionable activity” expressions were designed to contrast with each other, providing a highly visible method for the analyst to quickly determine the health status and the nature of the health data in view.

INTRODUCTION
To achieve optimal information visualization and visual analytic solutions [1, 2], rapidly and with less effort, we have designed and developed an open, adaptable, and extensible software development framework for producing visualization applications for analysts and decision makers in any common web browser. This framework—its design, API, and implementations—is called Aperture.

1.1 The Aperture Framework
In creating Aperture, the goal was to combine visual and analytic approaches, and create an expressive medium for programming graphics and visualization more efficiently. In addition, we wanted to enable easier transitions for developers without specialized expertise in graphics programming or visualization design.

The Aperture API provides a high-level vocabulary of visualization constructs, such as plots, bar and line series, indicators, nodes, and links, rather than low-level graphic primitives that must be chained together. This is a vocabulary more familiar to analysts, intended to help preserve context when assembling visualizations.

While combining primitive visual elements in novel ways provides one means of expressiveness, the approach in Aperture is to layer more intuitive and functional constructs in novel combinations. An extensible set of layer forms are provided, including an extension of the existing broad capabilities of the OpenLayers [3] open source toolkit to support dynamically interactive, geo-located visual entity representations. Chart, timeline, and network visual forms are also supported.

* [djonker, slangevin, pschretlen, ccanfield]@oculusinfo.com
to detailed event counts, charts by branch, and machine type distributions for a region or center; side-by-side comparison of selected regions or centers; time controls; and interactive vertical scaling of charts.

We also observed a pattern of after-hours maintenance in regions that were outside of business hours as of 2:00 PM BMT on February 2nd. This pattern included workstations that should have been powered down according to policy, and included multiple connections to machines. Our comparison view displayed a consistent pattern of after-hours maintenance of workstations, occurring in small numbers, but steady across all affected regions.

1.5 MC1.2

Building on the results observed for MC1.1, we were able to detect multiple anomalous events in the BankWorld dataset.

First, by adjusting the visualization to show activity over the entire 48-hour span, and filtering this view to show only red policy status events, we were able to detect a clear trend of policy status degradation over time, in both the severity and number of policy status deviations.

Next, a 48-hour regional view, filtered to chart only green “Questionable Activity,” showed gaps in reporting for Regions 8 and 9 (Figure 4). Using the same view, but filtered to show only Performance Issues, we determined that reports of fully consumed CPU were absent from Regions 6 and 46. These anomalies indicated that reporting services were compromised in these areas.

Next, we broadened the scope of our visualization to reveal the broader trend of off-hours maintenance activity over time. In the 48-hour view, we saw that after-hours workstation maintenance continued over the entire 48 hour reporting period. Since our analytics reported after-hours conditions with an activity code of greater than 1 (signaling an abnormal condition), we regarded the activity as suspicious, though it is possible this activity represents normal, automated maintenance after hours.

Last, we rapidly compared 24-hour views of the data for February 2nd and 3rd, by using the forward and back buttons in the browser, and with red policy status filtered out. This revealed two strange activity trends in DC5 and Region 25 on February 2nd. On February 3rd, activity is uniform across all regions and data centers. However, on February 2nd, during business hours, Region 25 activity decreases, while DC5 activity increases. This anomaly could indicate a facility-wide problem being addressed in DC5, while a regional issue affected operations in Region 25.

CONCLUSION

Aperture proved to be an effective framework for rapidly creating a tailored, informative, and easy-to-use visualization tool for the task posed in MC1. This serves as an example of using Aperture to practice “agile” visual analytics using “big data.” The lessons from this effort will be used to develop a foundation for agile analysis on extremely large datasets using visual analytics.

REFERENCES