BIG MARKETING: The simple way of browsing system logs to detect Network Events

Jorge Kuday Picoaga (*)
Universidad de Buenos Aires
Argentina

Diego Martin Cesario (§)
Universidad de Buenos Aires
Argentina

ABSTRACT

"If you can't explain it simply, you don't understand it well enough"

Albert Einstein

The widespread use and the exponential growth of information generated by technologies such as social networks, mobile devices and cloud computing as well as increasingly sophisticated attacks are dramatically altering the landscape of security, collection and analysis of data. Our proposal is to be focused on the creation and dissemination of a catalog of visualizations, built in Tableau, easy to understand by any audience, that has no prior knowledge of the domain. Mapping often represents an explanatory theory applied to the visual evidence. We might be accustomed to thinking about data as fixed values to be analyzed, but data is a moving target. How do we build representations of data that adjust to new values every second, hour, or week? This is a necessity because most data comes from the real world.

Keywords: Visual Analysis, Bubble chars, Treemap, Heatmap, Traffic logs, Vast Challenge, Domain, Tableau.

Index Terms: H.5.2 [Information Interfaces and Presentation]: User Interfaces – Interaction Systems;

1 INTRODUCTION

The domain provided to us was the traffic logs / status alerts of a business network. In considering the implications of the varied nature of the potential targets, one component of the response are to develop effective intelligence analysis methodologies for shows with simple plots both external threats and important events.

We as database managers, and our experience with daily deal with this type of information extremely dense and complex, dedicated to build reports to users without audit domain knowledge and operating systems, we place our main concern and objective, beyond the slogans meet the competition, building a set of views that summarizes and displays information from large volumes of data into simple views of what happen in a business network accurately and visually friendly.

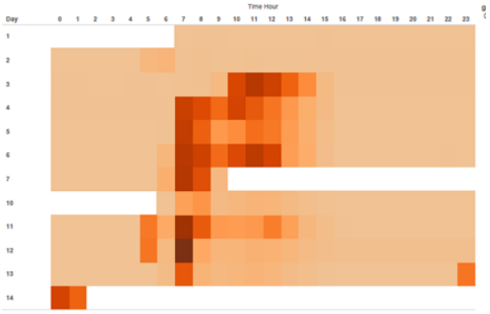
2 VISUALIZATIONS

In order to represent the most important events in the BB network, our experiments includes several visualizations to support analysis in getting an overview, finding trends and identifying suspicious events.

2.1 Visualizations for Health Monitoring

2.1.1 Heat maps

A Heat map, which is commonly used to display and compare Variables using a color intensity and position (in time in this case) was used to analyze the logs. It shows the moment and the strength of the event that takes place. The Heat map (Figure 1) was developed to show days and hours with the highest network activity.



Sum of gb (color) broken down by Time Hour vs. Day.

Figure 1: Heatmap is created to represent daily activity network in a 14 x 24 matrix, each cell represent a daily network activity hour.

2.1.2. BUBBLE CHARTS

Bubble plots have many benefits and one of them is to let you spot categories easily and compare them to the rest of the data shown just by looking at the size of the bubble.

In the figure 2, Hard disk percentage is represented for bubble plots. With some Tableau function animation we can see the evolution of the hard disk status and high percent used.

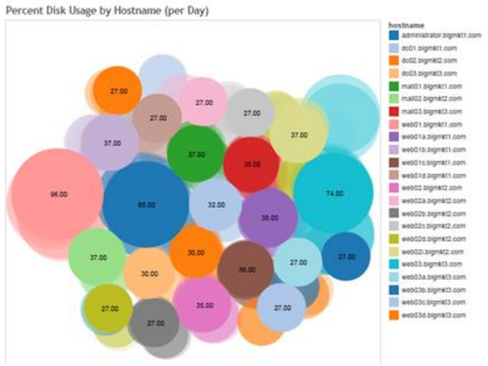


Figure 2: Bubble charts to visualize hard disk use in the network

2.2 Visualizations in order to show outliers

2.2.1 Treemaps Charts

Ben Shneiderman created Treemaps to display large numbers of values that exceed the number that could be displayed more simply and effectively using a bar graph.

In order to detect external attacks, inverted Treemaps were created; each small rectangle is an individual word in the Bbcontent column. The larger rectangles shows the words with minor frequency allow us observe the text outliers by text mining technic. A Treemap is called a space-filling display, because it takes full advantage of the available space. A Treemap displays parts of a whole and does so in a way that handles hierarchies. In our experiment, we can detect words that represent external attacks, virus and unusual reboots in the network. Rectangles within rectangles are used to separate the groups.

A Treemap is the best option for the implementation to be more thorough and reserved for large sets of values.

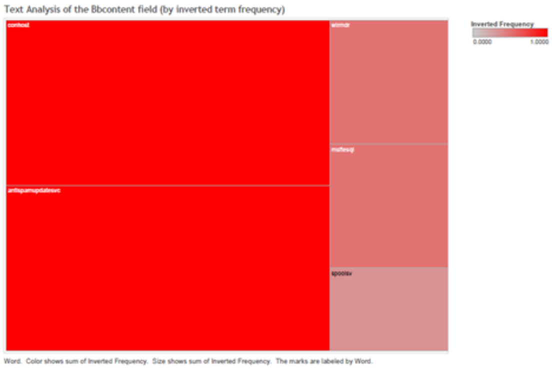
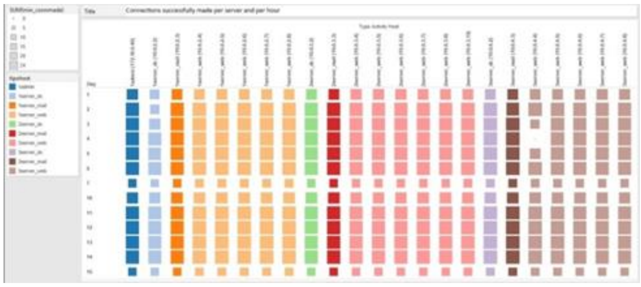


Fig 4. Inverted Treemaps charts to detect text ourliers in the network as virus or strange reboots time.

2.2.2 Simple plots by color

In the network description that reports the next paragraph: Organizationally, Big Marketing consists of three different branches, each with around 400 employees and its own web servers. Therefore the BigMktnetwork.txt 408 IP address that involves site 1 are reported, 407 IP address that involves site 3 are reported and suspiciously 308 that involves the site 2. Our theory was detect the 100 work station absence and to provide outliers information simple plot by color visualizations can be used. Basically a square representation by color is used and enhanced represent two variables (day, activity) using lines to represent 24-hour time-series data as shown in Figure 5.To analyze events Big Brother network health monitoring program reports 100 workstations status which was not declared in the BigMktnetwork.txt list.



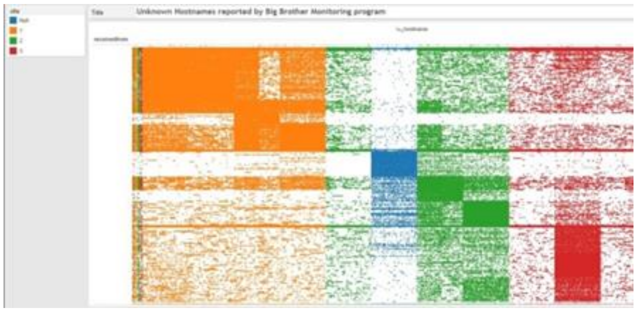


Fig 5. Simple plot with color to visualize outliers

3 Conclusion

In order to get the most important event occurred in our network we uses a combination of multiple simple visual representations.

The order of our catalog can be applied to control room situations to analyze and present analysis results in a scalable way. Through these multiple cross-linked visualizations several trends, patterns and suspicious events have been successfully identified in Big Marketing network. Effective data visualizations using Data Analysis intelligence technics is an important tool in the decision making process. It allows business decision makers examine large amounts of data, expose trends and issues efficiently, exchange ideas with key players and influence the decisions that will ultimately lead to success.

Mapping data by hand can be satisfying, yet is slow and tedious. So we usually employ the power of computation to speed things up. The increased speed enables us to work with much larger data sets of thousands or millions of values; what would have taken years of effort by hand can be mapped in a moment.

As network administrators of Big Marketing, it would be interesting to repeat our model two times per month with a graphical timeline to report interactively important issues in order to maintain informed to our network team.

Acknowledments

The authors wants to thank to Universidad de Buenos Aires for provide us the facilities in order to get our results test.

References

[1] Heatmaps http://en.wikipedia.org/wiki/Heat_map
[2] Edward Tufte, Beautiful Evidence 2006 Pag-41.
[3] Scott Murray O`Reilly -Intertactive data visualization for the web Pag-2.
[4] Ben Fry O`Reilly - Visualizing data Pag-18

* e-mail: georgepicoaga@gmail.com
§ e-mail: diegomcesario@gmail.com