MITSUBISHI ELECTRIC RESEARCH LABORATORIES
http://www.merl.com

Envisioning Grid Vulnerabilities: Multi-dimensional Visualization for Electrical Grid Planning

Romero, R.; Diez, D.; Wittenburg, K.; Diaz, P

TR2012-039   May 2012

Abstract

Electrical grid planning aims at optimizing the grid through the control of the performance and placement of electrical assets in order to minimize failures or vulnerabilities. With this purpose, grid planners carry out an initial stage of data exploration using a large volume of incident and equipment data collected over extensive time periods. In current practice these tasks are performed manually, which makes it very difficult to recognize patterns and gain insights into the data. In this paper, we propose a parallel multivariate visualization technique as a suitable approach for improving the existing practice. Based on the uage of an interactive visualization tool called BarExam, we demonstrate the feasibility of this visualization technique for displaying the dataset and present example insights that this visualization techniques can provide to grid planners.

International Working Converence on Advanced Visual Interfaces (AVI)

Envisioning Grid Vulnerabilities: Multi-dimensional Visualization for Electrical Grid Planning

Rosa Romero
Laboratorio DEI
Universidad Carlos III
Leganés (Madrid). Spain
+34 916249419
rmromero@inf.uc3m.es

Kent Wittenburg
Mitsubishi Electric Research Labs
201 Broadway
Cambridge, MA 01940 USA &
Universidad Carlos III Madrid
wittenburg@merl.com

David Díez
Laboratorio DEI
Universidad Carlos III
Leganés (Madrid). Spain
+34 916249499
david.diez@uc3m.es

Paloma Díaz
Laboratorio DEI
Universidad Carlos III
Leganés (Madrid). Spain
+34 916249456
pdp@inf.uc3m.es

ABSTRACT
Electrical grid planning aims at optimizing the grid through the control of the performance and placement of electrical assets in order to minimize failures or vulnerabilities. With this purpose, grid planners carry out an initial stage of data exploration using a large volume of incident and equipment data collected over extensive time periods. In current practice these tasks are performed manually, which makes it very difficult to recognize patterns and gain insights into the data. In this paper, we propose a parallel multivariate visualization technique as a suitable approach for improving the existing practice. Based on the usage of an interactive visualization tool called BarExam, we demonstrate the feasibility of this visualization technique for displaying the dataset and present example insights that this visualization technique can provide to grid planners.

Categories and Subject Descriptors
H.5.m [Information Interfaces and Presentation]

General Terms
Information Visualization, Design.

Keywords
Electrical grid planning, multi-dimensional visualization, bargrams, set-valued attributes

1. INTRODUCTION
An electrical grid is an interconnected network for delivering electricity from suppliers to consumers. It is controlled by a complex distributed system of sensors and devices that monitor both its electric and mechanical characteristics. The data gathered from this system is not only examined during operating tasks but are also utilized for offline trending and other analytical auditing. With the increasing complexity of the network, research and investments in electrical grid planning have been recognized as highly relevant by many countries [4]. This activity aims at optimizing the grid functioning through the control of the performance of electrical assets in order to identify failures or vulnerabilities. To achieve that, grid planners carry out a first stage of data exploration of large volumes of multivariate data such as historical data regarding equipment and past failures in the form of events. This data is spread out across multiple sources and has a number of different attributes. Examples of attributes are incident identifier, incident detection time, cause of the incident, and electrical assets affected by the incident. Conducting a semi-structured interview with two electric grid managers during a guided visit to an electric grid control center situated in Madrid, we observed that this first stage is currently performed manually through the application of simple data exploration strategies using spreadsheet tools (see Figure 1), sequential search through the entire dataset (row-by-row), and application of basic filters.

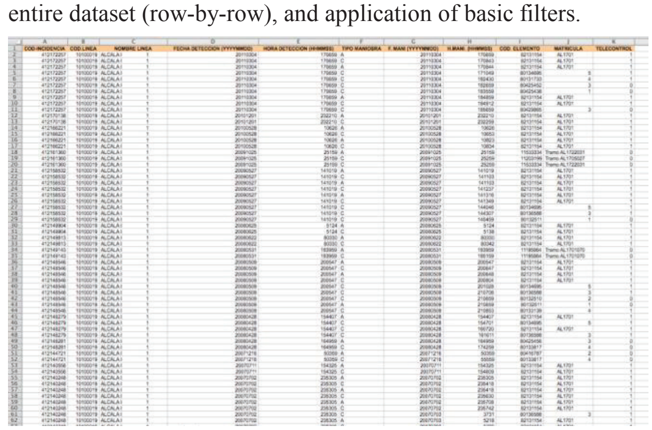

Figure 1. Extract of historical data records displayed in a spreadsheet tool

Since this current practice makes it very difficult to recognize patterns of grid vulnerabilities, we propose the application of a *parallel multivariate visualization technique* as a suitable approach that can help grid planners derive meaningful insights based on distributions and correlations of attribute values.

The paper is structured as follows. In section two we discuss related work on interactive information visualization methods for the electrical grid and other closely related operations domains. Section three presents the application of our proposal to real electrical grid data and the insights that this visualization method can afford. Finally, conclusions and future work are drawn in the last section.

2. RELATED WORK

The electrical grid is not the only networked infrastructure with distributed control in which its vulnerabilities must be analyzed. This section reviews the application of visualization research for supporting pattern extraction tasks not only within the electrical grid domain but also within other distributed-controlled infrastructures such as data networks. Within the security field, visualization has been given strong attention in the research community.

Due to the fact that electrical grids have a fixed number of nodes that are physically connected by transmission lines, *graph-based visualization techniques* [14,16] have been used for modeling the operations and physics of the electrical grid. The most significant result from these approaches is the ease with which natural electrical grid separation points can be identified if an uncontrolled islanding event occurs [16]. On the other hand, since a major event in the grid can be followed by a large number of alarms and subsequent events, *parallel coordinates* [11] in coordination with alarm lists have been applied to support the identifying of essential alarms. In the *parallel coordinates* display, each alarm attribute is represented by a parallel axis and each alarm is represented with a line crossing through the vertical axis at the alarm's corresponding attribute value. From this display, operators are allowed to explore alarm patterns and find out if there seems to be a correlation between some of the attributes, for example alarm type and time.

Since the number of security incidents reported per annum in data networks continues increasing [15], visualization research within the networking field has become a key for supporting offline intrusion detection tasks. These tasks involve the analysis of intrusion related data, which is stored in log files, in order to make appropriate decisions. Each of these incidents may involve multiple sites and within each site, it is highly likely that a large number of machines may fall under attack. Therefore, several visualization techniques have been proposed, which are mainly focused on supporting both visual pre-attentive processing and pattern detection tasks on these logs. Some of the most common techniques include the usage of *glyphs* [2] to represent nodes in a network; *color maps* [3] for differentiating between nodes and hosts; *histograms* [6] that usually represent the hosts and the period of time of the activities of each host; and *parallel coordinates* [1] to enable the user to identify activity patterns or sequence of packets transferred in a network.

The above summary shows that *parallel coordinates* have been the most popular visualization technique for supporting pattern extraction tasks in both domains. However, some limitations of this visualization technique have been acknowledged. They are most suited for numeric rather than categorical data, and some have claimed that over plotting is a major problem at larger scales [10]. Moreover, they do not handle set-valued attributes, which are a natural representation for incident data (e.g., the set of electrical assets affected per incident).

3. APPLYING VISUALIZATION TO ELECTRICAL GRID PLANNING

Our proposal involves *Parallel Interactive Bargrams* to support grid planners during the first exploratory stage. *Parallel Interactive Bargrams* have been successfully applied in systems that support consumer-based information exploration [17] and, more recently, for managing patent portfolios [18].

This visualization technique is derived by 'tipping over' the columns (bars) of the histogram and laying them end-to-end, ignoring any null bins. It uses a common representation for numeric values, *Booleans*, text, dates, time, and categorical data. Through *brushing techniques* with simple button pushing, a user can select items via their values in one dimension and those selected items' values are colored in all other dimensions, revealing correlations and other patterns. The working data sets can be quickly and easily filtered. Aggregation methods allow for scaling up to sets of tens of thousands of items. While the numbers of attributes do not scale up to more than dozens, this limitation is suitable for our domain task. Furthermore, since much data in the real world is naturally represented as sets, this technique supports the visualization of set-valued attributes.

In what follows, given the multi-source nature of this dataset, some data transformations necessary for preprocessing are presented. Afterwards, insights that this visualization tool can provide to grid planners are described.

Data Transformations

Since the required dataset is spread out across distributed databases, grid planners have to navigate across different spreadsheets in order to get a unified view of it. Unfortunately, this situation can provoke that grid planners get trapped in a phenomenon called *attentional tunneling* [7]. When people process information from multiple sources, they lock in certain aspects they are trying to process, and will inadvertently drop their scanning behavior. Aiming at overcoming this issue, we carried out a prior data transformation based on classifying such heterogeneous data under an integration schema. Grounded in expert knowledge, per each past grid failure, we defined a unified structure composed of three main aspects: (1) type of incident; (2) electrical assets affected; and (3) time range. Based on this integration schema, we created a database for enabling the application of *Parallel Interactive Bargrams* (see Figure 2).

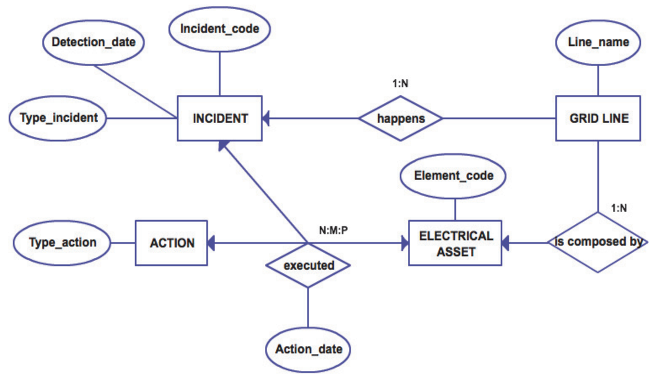

Figure 2. Entity-relationship diagram of the integration schema

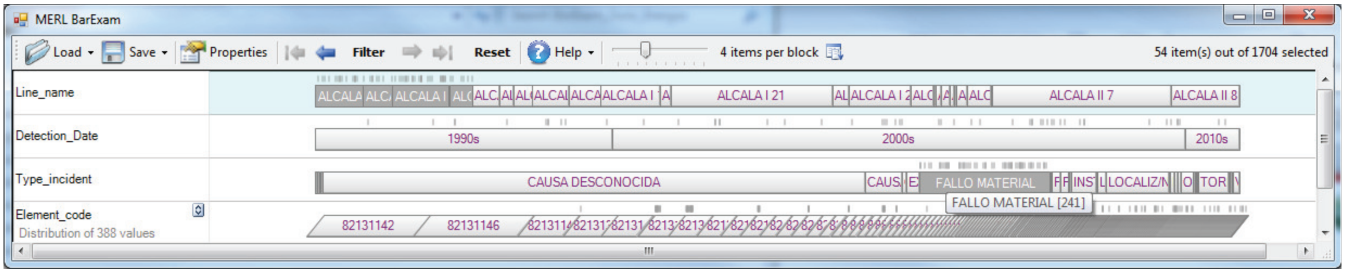

Figure 3. Screen shot of BarExam software. The values selected are colored in grey and set-valued dimensions such as Element_code are displayed as parallelograms.

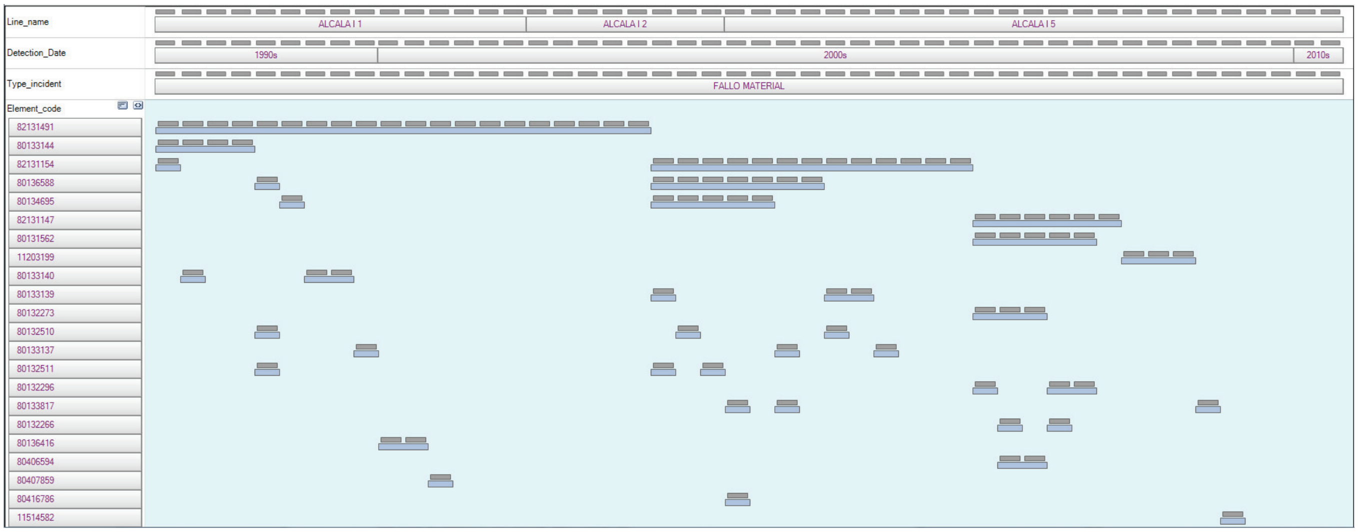

Figure 4. Screen shot of BarExam software. View is filtered following Figure 3 selections. Detailed view of the Element_code dimension shows co-occurrence through vertical alignment with the elements ordered by frequency.

Visualizing Grid Vulnerabilities

Here we present a use case of the application of the BarExam software [18] to this dataset that yields different insights for grid planners. This dataset consists of 1703 past grid failure incidents.

In the BarExam user interface, feature dimensions are presented as rectilinear *bargrams*, although in the case of set-valued dimensions they are presented as parallelograms (e.g. Element_code dimension in Figure 3). The item vectors above each *bargram* render individual incidents as *glyphs* that are used to indicate restrictions. The value bins in each *bargram* row are clickable; when they are selected they change to grey color. Value restrictions within dimension rows are treated as logical OR. On the other hand, value restrictions across rows are logical AND.

In Figure 3, the user selects four grid lines (row 1), and also a specific type of incident called *"Fallo Material"* (row 3). This type of incident is related to a cessation of normal operation such as power outages. As a result of this query, the uneven distribution of these incidents across electrical assets (row 4) is shown. Here is the first insight: a grid planner has realized that some assets are more affected than others according to the type of the incident. The user then filters the view and seeks more details on the Element_code dimension. BarExam provides the option of a more detailed attribute view that can consist of a normal horizontal histogram or a gapped histogram *Co-occurrence View* for set values such as is shown in Figure 4. This gapped histogram displays the distribution of incidents per electrical asset with the most frequently occurring at the top. Co-occurring assets appear underneath. Here we have an example of a second type of insight that grid planners can gain through this visual exploration: the co-occurrence of affected assets associated with incidents. The two most frequently affected assets do not co-occur most of the time, so the grid planner does not initially suspect a causal relationship.

Achieving such insights through existing practice involves defining several counters per dataset dimension, for instance, one incident counter per grid line. In such a way, determining the correlation among dimensions becomes a tedious task that mainly relies on the grid planner's experience and memory.

4. CONCLUSIONS AND FUTURE WORK

Our review of related work in the domains of electrical grids and networking security indicated a need for visualization approaches for supporting an initial data exploration stage. Trying to improve the current practice in terms of vulnerabilities detection, we have proposed the application of *Parallel Interactive Bargrams* as a suitable visualization technique that allows displaying fairly large amounts of historical data records regarding past events and failures. With easy-to-understand querying and filtering, users can visualize distributions and some relationships among past events and electrical assets. Since the required dataset is spread out across distributed databases, we have carried out a preprocessing step of data transformation based on classifying such heterogeneous data under an integration schema. To test the feasibility of this visualization technique, we have displayed this data with the BarExam software. After that, we have illustrated some insights that grid planners can gain through this visualization: (1) the understanding of the most affected electrical assets per type of incident; and, (2) the understanding of which electrical assets co-occur with others.

Further work will broaden our method's use to other datasets that grid planners would like to use to explore the first stages of their activity such as performance measures of electrical assets and periodic maintenance tasks. Our view is that parallel interactive bargrams can provide one part of an overall solution to grid planning, but we expect to combine them with other visualization and analytic methods. Once we have built a more complete solution, we expect to test our methods empirically.

5. REFERENCES

[1] Abdullah, K., Lee, C.P., Conti, G., Copeland, J.A. and Stasko, J. 2005. IDS rainStorm: visualizing IDS alarms. *IEEE Workshop on Visualization for Computer Security 2005 VizSEC 05* 1-10.

[2] Axelsson, S. 2003. Visualisation for Intrusion Detection: Hooking the Worm. Proceedings of the 8th European Symposium on Research in Computer Security (ESORICS 2003). Gjovik, Norway, Springer-Verlag.

[3] Ball, R., Fink, G., and North, C. 2004. Home-Centric Visualization of Network Traffic for Security Administration. Proceedings of the 2004 ACM Workshop on Visualization and Data Mining for Computer Security Washington, DC, USA, ACM Press.

[4] Chiaradonna, S., Giandomenico, F. and Lollini, P. Evaluation of Critical Infrastructures: Challenges and Viable Approaches. Architecting Dependable Systems V, Rogério Lemos, Felicita Giandomenico, Cristina Gacek, Henry Muccini, and Marlon Vieira (Eds.). Lecture Notes In Computer Science, Vol. 5135. Springer-Verlag, Berlin, Heidelberg 52-77. 2008.

[5] Colombe, J., Stephens, G. 2004. Statistical Profiling and Visualization for Detection of Malicious Insider Attacks on Computer Networks. Proceedings of the 2004 ACM Workshop on Visualization and Data Mining for Computer Security. Washington, DC, USA, ACM Press.

[6] Debar, H., Dacier, M., and Wespi.A. 1999. Towards a Taxonomy of Intrusion-Detection Systems. Computer Networks, 31(8):805–822.

[7] Endsley, M.R., Bolté, B., and Jones, D.G. *Designing for Situation Awareness.Design* 30-34 (Taylor & Francis: 2003)

[8] Erbacher, R. F., Garber, M. 2004. Fusion and Summarization of Behavior for Intrusion Detection Visualization. Proceedings of the IASTED International Converence On Visualization, Imaging, and Image Processing. Marbella, Spain

[9] Greitzer, F.L., Dauenhauer, P.M., Wierks, T. and Podmore, R. 2009. Human Factors Evaluation of Advanced Electric Power Grid Visualization Tools. Retrieved February 16, 2012, from Pacific Northwest National Laboratory: http://www.pnl.gov/coginformatics/media/pdf/EIOviz-greitzer-4-30_18386.pdf

[10] Kosara, R. 2010. Parallel Coordinates. Retrieved February 16, 2012: http://eagereyes.org/techniques/parallel-coordinates

[11] Mikkelsen, C., Johansson, J., and Rissanen, M. 2011. Interactive Information Visualization for Sensemaking in Power Grid Supervisory Systems. In Proceedings of the 2011 15th International Conference on Information Visualisation (IV '11). IEEE Computer Society, Washington, DC, USA, 119-126.

[12] Mitsubishi Electrics Research Laboratory. Retrieved February 16, 2012: http://www.merl.com/

[13] Muelder, C., Ma, K., Bartoletti, T. 2005. A Visualization Methodology for Characterization of Network Scans. Workshop on Visualization for Computer Security (VizSEC 2005)

[14] Overbye, T.J., Weber, J.D. 2000. Visualization of power system data. In Proceedings of the 33rd Annual International Conference on System Sciences.

[15] Symantec Global Internet Security Threat Report, 2010. Retrieved February 16, 2012, from Symantec Enterprise Security: http://eval.symantec.com/mktginfo/enterprise/white_papers/b-whitepaper_internet_security_threat_report_xv_04-2010.en-us.pdf

[16] Wong, P. C., Schneider, K., Mackey, P., Foote, H., Chin, G., Guttromson, R., and Thomas, J. 2009. A novel visualization technique for electric power grid analytics. IEEE Transactions on Visualization and Computer Graphics *15*, 410-423

[17] Wittenburg, K., Lanning, T., Heinrichs, M., and Stanton, M. 2001. Parallel bargrams for consumer-based information exploration and choice. Proceedings of the 14th annual ACM symposium on User interface software and technology (UIST '01). ACM, New York, NY, USA, 51-60.

[18] Wittenburg, K., Malizia, A., Lupo, L., and Pekhteryev, G. 2012. Visualizing Set-Valued Attributes in Parallel with Equal-Height Histograms. Proceedings of Advanced Visual Interfaces (AVI). Capri, Italy. ACM Press.