VAFLE: Visual analytics of firewall log events

Conference Paper *in* Proceedings of SPIE - The International Society for Optical Engineering · February 2014

DOI: 10.1117/12.2037790

4 authors:

Mohammad Ghoniem
Luxembourg Institute of Science and Technology (LIST)
48 PUBLICATIONS **787** CITATIONS
SEE PROFILE

Georgiy Shurkhovetskyy
University of Bonn
3 PUBLICATIONS **23** CITATIONS
SEE PROFILE

Ahmed Bahey
Nile University
2 PUBLICATIONS **13** CITATIONS
SEE PROFILE

Benoît Otjacques
Luxembourg Institute of Science and Technology (LIST)
18 PUBLICATIONS **82** CITATIONS
SEE PROFILE

VAFLE: Visual Analytics of Firewall Log Events

Mohammad Ghoniem[a], Georgiy Shurkhovetskyy[b], Ahmed Bahey[c] and Benoît Otjacques[a]

[a]CRP - Gabriel Lippmann, 41 rue du Brill, L-4422 Belvaux, Luxembourg
[b]M.S.A. University, Wahat Road, 6th of October City, Egypt
[c]Nile University, C.I.T School, 6th of October City, Egypt

ABSTRACT

In this work, we present VAFLE, an interactive network security visualization prototype for the analysis of firewall log events. Keeping it simple yet effective for analysts, we provide multiple coordinated interactive visualizations augmented with clustering capabilities customized to support anomaly detection and cyber situation awareness. We evaluate the usefulness of the prototype in a use case with network traffic datasets from previous VAST Challenges, illustrating its effectiveness at promoting fast and well-informed decisions. We explain how a security analyst may spot suspicious traffic using VAFLE. We further assess its usefulness through a qualitative evaluation involving network security experts, whose feedback is reported and discussed.

Keywords: cyber security, visual analytics, heatmap, clustering, firewall log, user study

1. INTRODUCTION

With the ever-growing amount of network traffic captured by firewalls and Intrusion Detection Systems (IDS) in corporate networks, the task of security analysts is becoming more challenging than ever. Scrutinizing millions of log entries each day to find anomalies requires more efficient tools to accomplish cyber situation awareness. For many years, visualization has been the best candidate to explore and communicate about large amounts of information. However, many security analysts have been working with fragmented tools including scripts and graphing capabilities without taking advantage of the potential offered by integrated visual analytics tools. This gap between the visual analytics and the security research communities has been coined as the "dichotomy of security visualization".[1]

In this work, we present VAFLE, a visual analytics prototype designed to achieve situation awareness and detect anomalies in the context of network security. We contribute a simple yet effective state of the art visual analytics system applied to firewall log events in particular, as well as customized clustering features that help automate the discovery of network activity patterns and enhance the scalability of the prototype. We evaluate the usefulness of the proposed combination of visualizations and clustering techniques in a use case using realistic datasets published by the "VAST challenge" in 2011 and 2012, and illustrate its effectiveness at promoting fast and well-informed decisions by network analysts. We further set up a qualitative evaluation where we ran VAFLE by several network security experts, whose feedback is reported and discussed in the light of the latest research works on visualization evaluations.

In the next section of this paper, we overview related work in network security visualization. Then we discuss the needs of network security analysts (section 3) and the types of data they usually investigate (section 4). In section 5 we present the main features of the VAFLE system. In section 6, we illustrate its usefulness in achieving typical network security analysis through a realistic use case. In the penultimate section, we describe a qualitative user evaluation conducted with network security experts, then we report and discuss their feedback. Lastly, we discuss the limitations of the present work and future developments.

Further author information: (Send correspondence to Mohammad Ghoniem)
Mohammad Ghoniem: E-mail: ghoniem@lippmann.lu, Telephone: +352 47 02 61 623
Georgiy Shurkhovetskyy: E-mail: shurkhovetskyy@gmail.com
Ahmed Bahey: Email: ahmed.bahey@nileu.edu.eg
Benoît Otjacques: Email: otjacque@lippmann.lu

2. RELATED WORK

Although security visualization is a young research field, it has received much attention in the last few years. Several tools have been designed to help network security analysts gain insight into network traffic and detect anomalies. A recent survey by Shiravi et al.[2] provides a comprehensive list as well as a scenario-based taxonomy of network security visualization tools. Authors note that most network security visualization systems are designed for only one type of data, for example, raw packet captures, IDS alerts or firewall logs etc. By focusing on firewall log data, VAFLE also falls in this category.

The visualization of large IP spaces is a core challenge in network security visualization as large computer networks may contain tens of thousands of hosts.[3] The problem is aggravated as the IPv6 protocol is rolled out, making the available number of pixels on a computer monitor too small to display so many hosts.[2] In this section, we only discuss tools that use a heatmap or grid-based representation of network security data as they are visually similar to VAFLE. Host×host, host×time and time×time grids have been used in previous work. For instance, TNV[4] uses a host×time grid where hosts are sorted in IP order and cell color encodes the number of IDS alerts fired concerning any pair of host and time slice. Links between sending and receiving hosts within a single time slice are plotted as straight lines on top of the grid, resulting in visual clutter. A supplemental parallel coordinates view displays the links between source and destination ports and provides various filters. Hypothetical scenarios show analysts' ability to use TNV in order to detect SNMP and portscan attacks and learn normal network traffic patterns. VISUAL[5] displays internal hosts on a host×host grid. External hosts are displayed as squares laid out around the grid whose size is proportional to the traffic they engage in. Links between internal and external hosts are displayed as straight lines that fan out/in from the grid cells. Like TNV, line overplotting results in visual clutter. A slider is used to play back network activity interactively. Tracking host activity over time through animated fan ins/outs remains difficult. Port information is displayed as horizontal lines on top of each square/external host, adding to the complexity of the view. VISUAL is said to be suitable for a mid-sized network with about 2,500 internal hosts connected to approximately 10,000 external hosts. Contrasting with VAFLE, neither TNV nor VISUAL have any multi-level exploration or clustering capabilities, limiting their ability to scale to larger networks or automate the discovery of network traffic patterns.

In the IDS Rainstorm system, a manifold host×time grid is used to visualize the equivalent of 2.5 class B networks on one screen, where each row aggregates 20 hosts. Color hues encode IDS alert severity. A pixel is colored according to the most severe related alert. Internal IPs are laid out on the left side of the grid in IP order, while external IPs are laid out on the right. Arrows connect alerts to the related external hosts. A secondary window provides a magnified view of a user defined network range. The tool also supports filtering and details on demand interactions. However, it lacks a visualization for port information. In contrast, VAFLE provides host×time and port×time visualizations. Scalability is achieved through clustering and multi-level navigation in aggregate views powered by using a database backend. We tolerate horizontal and vertical scrolling when detailed views cannot fit in available screen real-estate. Moreover, we use VAFLE to analyze firewall rather than IDS logs and, hence, address a different set of questions as detailed in section 3.

Hierarchical representations of IP space have also been used to display aggregate network data. A Treemap visualization is used to map network communications in NFlowVis,[6] NVisionIP[7] and HNMap.[8] Internal hosts are mapped to rectangles whose size and color encode attributes of interest. In NFlowVis, the treemap represents hosts related to attackers during a time frame. External hosts are plotted at the borders of the treemap and connected to internal hosts through splines. NVisionIP uses animation to play host activity over time, whereby changes in activity levels are perceived as color change, which makes it difficult to track the activity of multiple hosts simultaneously. Following a grid-based approach, ClockView[9] depicts host activity (netflows) over a 24-hour time window using a clock glyph, where hourly traffic is mapped to a colored wedge around a 24-hour clock. Color gradients encode either sheer traffic figures or differential values over several consecutive days. Such glyphs are then laid out in a 256×256 grid setup, accommodating class B networks, provided that the user scrolls horizontally and vertically over multiple screens. Glyphs can be ordered in several ways e.g. in subnet vs. host order like in NVisionIP, or in descending traffic level order. In subnet order, clock-wedge orientation across hosts in the same subnet allows the user to spot synchronous traffic trends. Several filtering options are provided. Inter-computer links are displayed on demand as semi-transparent straight lines on top of the matrix, creating visual clutter. For a selected pair of hosts, a matrix view displays source and destination port data.

A time×time grid visualization of network traffic was presented by Lamagna.[10] Both dimensions are used to represent time but at different granularities: columns represent hours of the day and rows represent minutes of the hour. Color intensity encodes the frequency of network events, while different color hues are used for different data sources. Events from IDS logs are laid over firewall data, resulting in occlusion problems. Data is preprocessed using perl scripts, which limits the scalability of the system.

A lot more tools are available in the literature. Some consider summarized information of network traffic. For instance, PortVis[11] displays summarized information about port activity as a pixel grid. Each pixel represents a port and color encodes any numeric value associated with the port such as the number of unique source IPs using it. Other tools such as NFlowVis use multiple data sources (netflows + IDS alerts) along with multiple coordinated views to enable quick visual insights into communication patterns. TVi[12] is one of few systems integrating data mining and visualization for the purpose of network forensic analysis. It combines machine learning with a matrix visualization of netflows. The user can set the display axes to source/destination IP or port. Cell color encodes flow intensity. Scalability to large datasets is achieved using Principal Component Analysis (PCA) backed by a RDBMS. In VAFLE, clustering is used to discriminate anomalous behavior and combined with a RDBMS to improve scalability. VizSec 2004[13] features many tools combining visualization and data mining for computer security purposes.

3. NETWORK SECURITY ANALYSIS

Network security analysts explore and analyze network traffic data to ensure system security. There are three categories of activities they typically perform: reporting, historical analysis, and real-time monitoring.[1] Reporting involves presenting information through reports, whether textual or graphical. Historical analysis involves analyzing data from the past for investigation, understanding data trends, finding anomalies without prior knowledge, etc. Real-time monitoring involves examining the current security state of a network to accomplish situation awareness. In this paper, we present an interactive visualization system for firewall log data to ease anomaly detection and achieve situation awareness. Following the information seeking mantra,[14] the analyst starts by getting an overview on the available network data answering questions like: "who are the top talkers?", "which network services are used?", and "which computers communicate with each other?"[1] Then, he may drill down into more details and focus on specific hosts or services running or any combinations of hosts and services that may have caught his attention.

4. DATA SOURCES

Among many network data sources available for analysis, firewall and IDS logs play a significant role in the forensic analysis of network traffic.[15] Sitting at the interface between a network and the outer world, firewall devices are a significant source of raw information that are invaluable in building cyber situational awareness. In fact, depending on their configuration, firewalls may log all attempted connections both incoming and outgoing. Typically, a firewall log contains information about the time, source and destination IP addresses and ports, the protocol used, a severity level indicator and whether the connection was allowed or denied.

On the other hand, IDS systems analyze network traffic, inbound and outbound, looking for patterns of common computer attacks. When such patterns are detected, alerts are fired requiring attention from the network security administrators. In contrast, a firewall silently logs connection information. IDS systems operate using two main approaches: signature-based detection or statistical anomaly-based detection. When tuned correctly, IDS logs present a significant source for a more focused view of threats. However, IDS systems generate so many alerts that it is impossible for analysts to investigate each and every one. Conti et al.[3] provide an insightful discussion of the information overload problem encountered by analysts dealing with IDS alerts. Considering that only a small share of alerts correspond to real threats and that threat detection heuristics generate a lot of false positives and false negatives[2] due to their sensitivity to noise, network security analysts cannot rely solely on the analysis of IDS logs. When they need to trace an IDS alert, analysts will eventually look up the related firewall log chunk in order to get the facts about network traffic. Likewise, when a real threat is wrongly discarded as a false negative, they stand a chance to pick up its trail by inspecting firewall logs.

Other sources of information such as operating systems logs and application logs may also help in cross-checking hypotheses and identifying the exploits leading to a security breach.[15] In VAFLE, we attempt to fulfill the security analyst needs listed earlier in section 3 by providing interactive visual analytics capabilities for the exploration of firewall log records allowing the analyst to use informative multiple coordinated multi-level views, apply filters and clustering on data and get details on demand.

5. INTERACTIVE VISUALIZATION

VAFLE is built using the Information Visualization Toolkit (IVTK) v0.10[16] which includes many well-known visualizations for the display of trees, graphs and time-series. It also supports many interaction mechanisms such as fisheye lens magnification,[17] excentric labeling[18] and dynamic queries.[19] In VAFLE, we extend it by building custom heatmap views and magic lens interaction and by implementing several clustering techniques tailored for the analysis of firewall logs. Scalability is achieved by relying on a modern RDBMS and is further enhanced through clustering.

5.1 Visualization of Top Talkers and Top Services

In order to capture the activity of multiple entities over time, in particular top talkers and top services, we build host×time (Figure 1) and port×time (Figure 2) heatmap grids[20] of firewall log data. The heatmap provides a clutter free overview of the traffic as recorded by the firewall. For example, in Figure 1 (top half), the columns stand for source IPs while the rows represent half-hour time slices across a 2-day observation period. Cell color encodes the number of connections established for any pair of source IP and half hour time slice. Since the data is sequential, a color gradient is used. The darker the color, the greater the traffic. Color equalization is used to mitigate the effect of outlier values. In the heatmap view, periodic activity patterns appear as parallel horizontal stripes as in the right half of Figure 3c. When columns are sorted in IP order, subnet activity patterns may be seen as vertical stripes as in the top half of Figure 1.

Heatmap grids offer the advantage of being easy to understand by untrained users. They have been used in many disciplines such as software visualization,[21,22] bioinformatics[23] and social sciences[20,24] and are therefore well-understood. Their semantics depend on the parameters mapped to rows and columns. Their ability to reveal patterns depends on row and column order.[24–26] Many systems described in the related work section use IP order so that hosts in the same subnet appear next to each other. It also seems suitable to sort ports in the ascending order in order to ease port lookup. Depending on the task, other sorting orders are also possible.[9] When rows are mapped to time, they will be sorted in the chronological order by default. By doing so, we have managed to detect temporal and structural patterns in network traffic as discussed in the use case (section 6).

In VAFLE, heatmap rows are mapped to time slices whose granularity may be set by the user (30 minutes by default). Similar heatmap views may be built for destination IPs instead of source IPs. In Figure 2, the heatmap represents port activity over time. The dark vertical stripes correspond to ports 53, 88, 135, 137–139, 389 and 445 all of which are often used by trojans and worms to spread across LANs. In this figure, these network services display much higher traffic than they would normally and much higher than other services. Moreover, a dark horizontal band reveals the presence of massive traffic on most ports taking place on April 14th from 11 AM to 2:30 PM. Obviously, this leads the analyst to investigate more in-depth a portscan attack scenario.

We also visualize top talkers and top services using classical time series line charts, as in the bottom half of Figure 1, where the activity of each host or port of interest is plotted as a spline with time flowing from left to right while the y-axis maps traffic levels. In Figure 4a, the line chart displays network traffic of all hosts over 2 days. Each curve stands for one host. Despite the clutter due to curve overplotting, the analyst can immediately spot two outliers responsible for significantly higher traffic than most other hosts. At the bottom, multiple curves seem to be bundled, rising and sinking together synchronously. However, this phenomenon is obscured by curve overplotting, which the analyst could ease to some extent using the IP filtering feature provided in VAFLE as in Figure 1. The analyst may then enable scale fitting to adjust the y-axis scale to the filtered data. Such coordinated host activity appears glaringly as dark horizontal stripes in Figure 3c, suggesting a botnet attack.

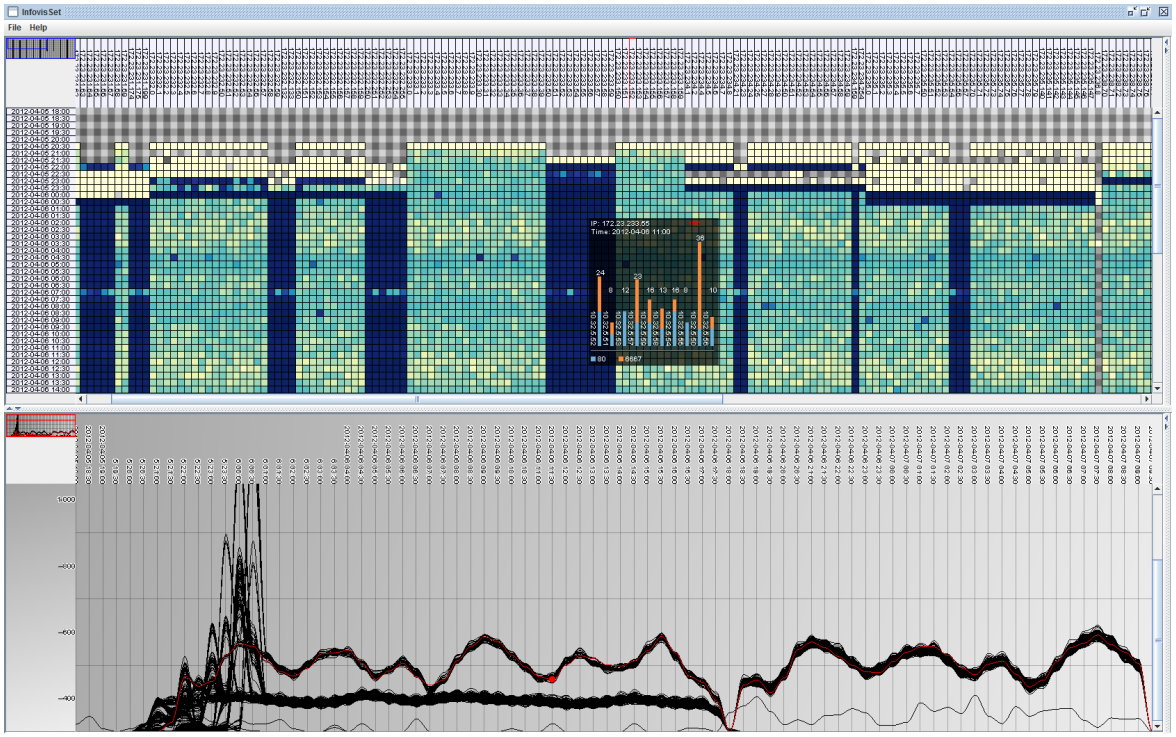


Figure 1: VAFLE main window showing two coordinated views of host activity over time as, at the top, a heatmap with a magic lens and at the bottom, as a line chart highlighting in red the same host and point in time selected on the heatmap.

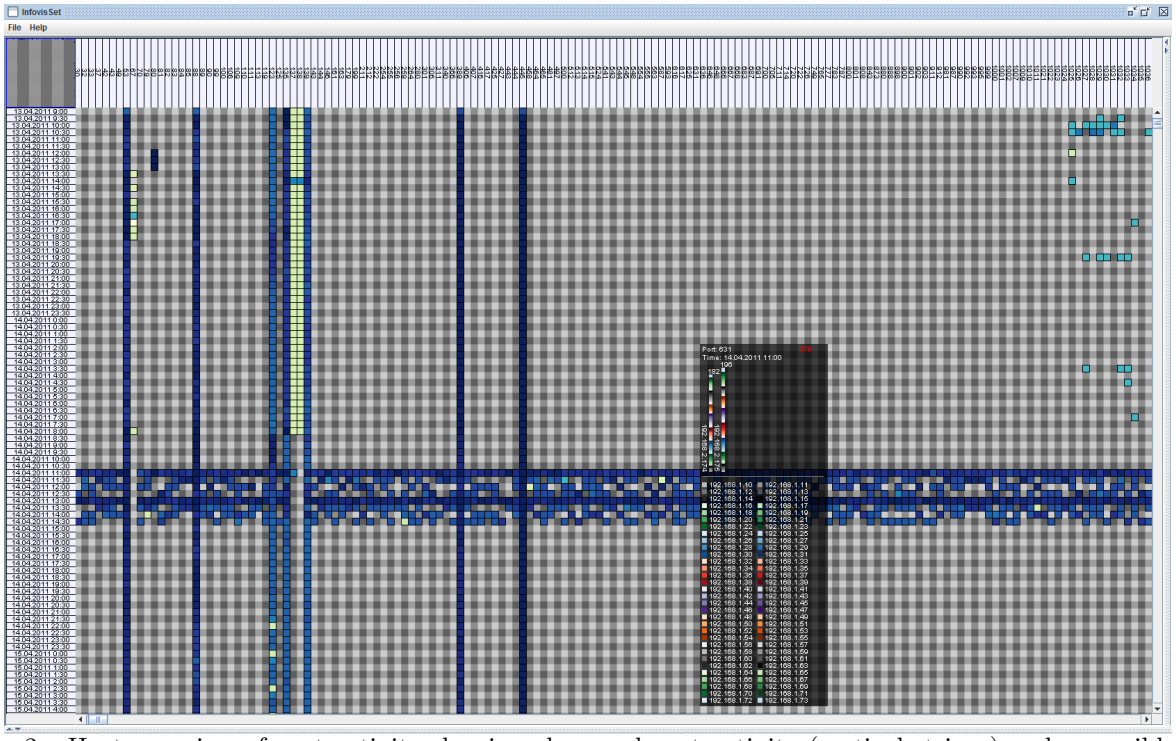


Figure 2: Heatmap view of port activity showing abnormal port activity (vertical stripes) and a possible port scan attack materialized by the horizontal blue band. The magic lens provides detail in context information about the attack.

5.2 View Coordination

VAFLE implements linking and brushing interactions allowing the analyst to take advantage of the strengths of both heatmaps and line charts and work around their weaknesses. When visual clutter is an issue, the analyst can move to the heatmap view. The sequential color coding of heatmaps allows the comparison of cell activity relative to each other and proves effective at showing patterns like in Figures 3c and 4c. When needed, exact connection counts can be accessed interactively using excentric labeling[18] on the heatmap view or by looking at the linked line chart where values can be gauged rapidly on the y-axis. VAFLE currently supports coordinated selection and filtering across the heatmap and line chart views, as highlighted in red in Figure 1. Selecting one host/port in the heatmap highlights the corresponding curve in the line chart and vice versa. Similarly, once a cell has been selected in the heatmap, a big colored dot is displayed on the corresponding curve and time slice for easier identification. Figure 1 shows an example of linking and brushing in VAFLE where a host and time slice of interest are highlighted in red. Likewise, filtering a set of columns in the heatmap through dynamic queries impacts the corresponding curves in the line chart view and vice versa. A magic lens shows the port activity breakdown of that host on the heatmap view at a user selected time slice in a focus+context fashion. Filtering certain IPs out simply removes the corresponding columns from the heatmap. However in the line chart filtering reduces the clutter significantly and helps the user inspect the behavior of specific hosts in detail. Instead of looking at overall aggregate figures, the analyst could filter the host×time visualizations to retain the connections made on any combination of ports and/or specific destination IPs only resetting the colors in the heatmap accordingly. Thus, abnormal activity regarding a specific port can easily be checked. For example, we found out that the periodic component in the botnet attack was due to web traffic, where an initial mix of web and IRC connections was first found. Such flexibility increases the user's ability to make and validate hypotheses.

5.3 Clustering

Towards an automated detection of traffic patterns, VAFLE supports two clustering algorithms: k-means and unnormalized spectral clustering.[27,28] The default number of clusters generated in VAFLE is $k \approx \sqrt{\frac{n}{2}}$ where n is the number of items in the dataset.[29] However, the analyst may specify the desired number of clusters through the GUI. The similarity metric underlying the clustering algorithms may be configured to use any available attribute (feature) of firewall log records. When investigating network traffic from a host centric point of view, an analyst may wish to discriminate groups of hosts with similar behavior patterns over time. We use a metric based on unnormalized cosine similarity[30] defined in the following equation:

$$1 - \frac{a \cdot b}{\|a\|\|b\|} = 1 - \frac{\sum_{i=1}^{n} a_i \times b_i}{\sqrt{\sum_{i=1}^{n} (a_i)^2} \times \sqrt{\sum_{i=1}^{n} (b_i)^2}}$$

where a and b represent columns in the matrix, n is the number of rows while each a_i represents the number of connections initiated by host a in time slice i, all of which together constitute vector a, whose angle is calculated with respect to host b. Hence, connection counts in all time slices contribute in host similarity computations. Yet another option consists in using the maximum number of connections per host per time slice as a similarity metric. This helps separate the top talkers from less vocal hosts, making the detection of anomalous activity more likely.

Another way of differentiating hosts consists in profiling the network services they use. More specifically, we propose the maximum number of distinct destination ports used by each host per time slice as a similarity metric. Clustering based on this attribute reveals groups of hosts having radically different behaviors on the network. Instead of including the whole available history, the analyst may also select a specific time interval (sub-vector) to be taken into account while clustering, combined with any of the aforementioned parameters. Search-by-example queries can typically be supported by this approach: when the user spots a suspicious event unfolding over a certain time interval involving a set of hosts or network services it becomes possible to find all hosts with a similar time-constrained behavior.

In VAFLE, clustering and filtering can be combined in an iterative visual analysis process. The analyst may first apply various filters on the traffic, which impacts the heatmap and line chart views. Then the filtered data

can be clustered. Various combinations of filters and clustering metrics would therefore provide various points of view on the traffic, thus empowering the analyst to achieve many different analyses depending on unforeseen needs. The cluster hierarchy makes it possible for the analyst to look at the data at multiple levels of detail, by expanding or folding any cluster of interest through direct manipulation of the visualization. Depending on the task at hand, the analyst may rather concentrate on clusters having few items (typically outliers, rare events) while taking a cursory look at highly populated clusters where item behavior is deemed "normal". Helping close the semantic gap between clustering algorithms and subject domain experts, VAFLE supports interactive post-editing of clusters by direct manipulation on the heatmap view. All cluster manipulations (fold, unfold, merge) are coordinated across views. A thorough motivation and discussion of interactive post-editing of cluster hierarchies can be found in recent literature.[31] This functionality can be seen as an intuitive alternative to tuning the similarity threshold of the clustering algorithm through the GUI and having to re-cluster the dataset all over again. Finally, the analyst may save the hierarchy of clusters for future discussions with collaborators.

5.4 Magic Lens Interaction

After getting a sense of the security state through the heatmap or time series overview, the analyst may need to obtain detailed information about suspicious phenomena related specific hosts or services. For this purpose, we implement a magic lens interaction[32] mechanism for both heatmaps and line charts displaying a stacked histogram on demand showing traffic breakdown. For instance, once the mouse rolls over a pair of source IP and time slice, VAFLE overlays a stacked histogram where bars stand for destination hosts. Each bar is further broken into stacked rectangles corresponding to the respective ports activity between the source and destination IPs. At one glance, the analyst can see in a context+detail fashion how many destination hosts are connected to the selected source IP and the related traffic distribution across various ports during any particular time interval. For example, in Figure 4a, the magic lens looks up detailed information about traffic originating from IP 172.23.235.51 between mid-night and half past mid-night featuring a mix of web traffic (light blue rectangles) and a majority of IRC traffic (rectangles in orange) directed to nine external hosts with IPs in the range 10.32.5.x. Whether such traffic is legit or requires further investigation is left to the analyst's appreciation. Likewise, looking at overall port activity over time in Figure 2, the analyst may point at any cell in the heatmap and use the magic lens to inspect traffic distribution among source and destination IP pairs in the selected time slice. In this figure, histogram bars correspond to source hosts while color encodes distinct destination hosts. The analyst spots two hosts with IP addresses 192.168.2.174 and 175 connecting to virtually every host in the subnet 192.168.1.x. In fact, probing the horizontal blue band in Figure 2 using the magic lens reveals that these same two internal hosts have been port scanning all hosts in subnet 192.168.1.x for three and a half hours. Lastly, when the user points to a cluster rather than a single host or port, the magic lens shows the activity pattern of the cluster centroid.

6. USE CASE

In order to test the usefulness of VAFLE against realistic data, we used the datasets published by the VAST Challenge in 2011 and 2012. The VAST 2012 Mini Challenge 2 consists in finding the five most noteworthy events occurring in the network traffic and identifying the root causes of abnormal computer behavior. Raw firewall and IDS logs are provided. The network policy of use and network topology are also described. Complaints from user about rogue security software running on their computers and constant hard drive I/O operations have been filed. The network traffic at hand involves approximately 5,000 hosts with over 46,500 distinct one-to-one links between them. The firewall log covers two days and contains around 23 million events. Important fields available in any firewall log format are source and destination IP and port information, along with event timestamps. Using a database backend, the data is aggregated counting the number of hits for every triplet (source IP, destination IP, destination port) in every time slice throughout the two days. These detailed values are displayed on demand in the magic lens. They are further aggregated based on source IP to generate the overview, making it possible to appreciate "how loud a computer is talking" over the network as in Figure 4a, or based on destination ports making it possible to overview top network services in use as in Figure 2. The time granularity defaults to 30-minute intervals but may be defined by the user interactively.

To start with, we achieve situation awareness by determining the top talkers and the top services throughout the two days of observation. To do so, the firewall data is loaded into both the heatmap and line chart views. The

detailed host×time heatmap features 5,000 columns and 96 rows. By clustering hosts based on their temporal activity profile (the whole column vector is used for similarity computations), the analyst can search the data for activity patterns. The clustering step generates a coarser heatmap having around 15 columns and 96 rows, that fits easily in one screen as in Figure 3a. The two left-most clusters show very distinct patterns. Wishing to get a close-up view, the analyst expands them interactively and gets a refined heatmap as in Figure 3c. The identification of horizontal stripes reveals a periodic traffic pattern across a cluster of 600 hosts approximately.

Alternately, the analyst may choose to cluster the hosts based on the maximum number of outgoing connections per time slice. A dark cluster throughout the two days appears on the right side of the heatmap in Figure 3b. It contains two hosts only (172.23.0.132 and 172.23.252.10) corresponding to the top talkers. Their traffic volume is much higher than any other host in the network as can be seen in the line chart view in Figure 4a. Internal hosts are not expected to initiate such a high number of connections which makes their activity highly suspicious. Investigating the behavior of these hosts more closely, the analyst can interactively get more details about their traffic using the magic lens. The two top talkers partake in a lot of web traffic directed towards external hosts in the range 10.32.5.x. They also initiate IRC connections (port 6667) to these destinations. IRC is often used for remote command execution and helps orchestrate botnet attacks from zombie computers. Moreover, since the corporate policy forbids IRC, its being used by the top talker hosts triggers further investigation about IRC traffic. Through port filtering, the analyst is able to identify all hosts engaged in IRC traffic. Then, applying clustering to these hosts reveals a variety of usage patterns of IRC. It is clear now that there are many hosts involved in the forbidden IRC traffic. Removing the filter on the IRC traffic lets the analyst look at all services used by these hosts. In particular, the analyst could observe the pattern seen in Figure 4c where a high web usage is confined among a few hosts in the first day (see the dark vertical stripes) and spreads to the whole range of hosts in the second day. Both this web traffic and the IRC traffic identified earlier are directed towards the same external hosts in the range 10.32.5.x. At this point, the analyst may have picked up the trail of a data ex-filtration attack.

Further scrutinizing these suspicious hosts, the analyst discovers that many of them have the same temporal activity pattern rising and sinking synchronously, as can be seen in Figure 4b. The magic lens also shows a scarce use of suspicious ports: 21, 22 and 6667, unseen with other hosts. Checking out inbound traffic coming from external hosts in the 10.32.5.x range reveals that they also targeted the corporate firewall at IP 10.32.0.1 and use many different ports every single time slice for such traffic. This leads the analyst to suspect a portscan attack on the firewall. Clustering the traffic based on the maximum number of distinct ports used per host per time slice separates the hosts into a handful of highly populated clusters with hosts using only 1 to 3 ports, and 2 small clusters, seen on the right in Figure 3d. Hosts in these two clusters engage in low traffic volumes, hence the dominant light yellow columns, but they use as many as 64 and 110 distinct ports per time slice on average, as shown by a very colorful magic lens and the detailed textual information panel on the right.

Examining the ports heatmap in Figure 2, multiple dark vertical stripes reveal consistently high traffic on certain ports across time. Outstanding activity can be spotted on port 80/Web and port 6667/IRC throughout the two days among other ports. Ports 21/FTP and 22/SSH feature shorter dark vertical stripes. Prohibited by the corporate policy, attempted FTP traffic requires further investigation about the source hosts even if the traffic was denied by the firewall. However, SSH connections were successfully established between internal hosts and multiple external hosts in the range 10.32.5.x, potentially enabling the latter to execute commands remotely on local hosts and to drop malware e.g. worms and trojans inside the network as first step towards data-ex-filtration. A more consistent corporate policy concerning file transfers should have denied both FTP and SSH traffic. Applying the magic lens to inspect ports usage in the port activity heatmap/line chart allows the analyst to pick up the trail of the aforementioned portscan attack on the firewall as well as the internal and external hosts involved in this traffic as detailed earlier.

Beyond the VAST 2012 mini challenge 2 for which it was built, VAFLE can handle any other firewall logs recording source IP, source port, destination IP and destination port of all network communications. Thus, we managed to use VAFLE to carry out a cursory investigation of the VAST 2011 Mini Challenge too. Without prior knowledge of the dataset, we were able to observe vulnerable ports being probed (e.g. network share ports). We were also able to quickly spot a portscan attack and a likely Distributed Denial of Service (DDoS) attack running over 4 hours and a half in a row. We skip further details in the benefit of space.

In this section, we have shown how a combination of multiple coordinated interactive visualizations and a variety of user-tunable clustering metrics may help network security analysts sift effectively through large volumes of firewall log data and pick up the trail of potential security threats. VAFLE has been tested against the VAST 2011 and 2012 Challenge datasets and allowed us to make interesting findings without prior knowledge of the data. These findings include what may qualify as botnet attacks, portscan attacks, DDoS attacks, and data ex-filtration attacks. The full identification of the root causes of those issues may yet require the fusion and analysis of complementary data sources such as packet captures, IDS alerts, application logs and the ability to check back firewall logs which contain the raw facts about network traffic. In order to further prove the validity of our approach in general, an expert evaluation of VAFLE was in order.

7. QUALITATIVE EVALUATION

The evaluation of visualization systems has received an increasing attention from the scientific community in the past few years. Many authors recognize the difficulty to perform user studies in visualization.[33–36] Since VAFLE was initially designed for the VAST 2012 challenge, without access to network security experts at that time, it appeared necessary that a summative evaluation of the system determined its usefulness from a security analyst standpoint and whether this audience would be willing to test it on their own data should an opportunity arise. In the scenario-based taxonomy provided by Lam et al.,[34] our user study falls under the evaluation of "User Experience" category. Following their recommendations, we chose to conduct a qualitative user study with few expert users asking them questions related to the usefulness of system features, whether features were missing or needed to be reworked and whether they would consider adopting it.

7.1 Experimental Setup and Procedure

We prepared a structured demo of the system and a list of features that we wanted to showcase to the participants. We also prepared a questionnaire divided in two parts. Firstly, closed-ended questions about the usefulness of system features where the expert is expected to provide a rating on a five-point Likert scale. Secondly, closed-ended questions aiming to assess the appropriateness of design choices made in VAFLE since no expert users were available to inform early design stages. Both sets of questions are summarized in Tables 1 and 2.

We conducted a qualitative evaluation to assess the VAFLE system by network security analysts. Three experts from the network security industry participated in the evaluation. All of them held a B.Sc. in Computer Science. The first expert, referred to as P1 in the sequel, is a senior security engineer with 8 years of experience in the field. The second, P2, is a team leader in information security with 5 years of experience. The third, P3, is a security technical support engineer with 2 years of experience, he holds an M.Sc. in Information Security. None of them has previously been trained in data visualization. All of them spend a significant share of their time fighting against network threats which involves examining firewall logs on a daily basis.

Experts were interviewed individually in their respective offices by two interviewers, one was running the interview and taking notes while the other was exclusively taking notes. Experts participated on a voluntary basis and no compensation was offered for their participation. Each interview lasted from one to two hours. We first discussed with each expert his regular duties and tasks, what approaches and tools are used at his organization to perform them. We then described VAFLE, its features, capabilities, different usage scenarios and ran a short demonstration. The expert was then invited to try VAFLE himself and make comments. At the end, the expert filled the questionnaire.

Because the demo was run on a commodity notebook computer, the dataset used in the evaluation was a subset of the 2012 VAST challenge data including a day and a half worth of firewall log for a network of 1,000 hosts displayed at a half hour time granularity. The data was pre-clustered: during the showcase the clusters appeared instantly. However, experts were informed that clustering takes about 4 seconds for the data at hand, and grows linearly with data size.

7.2 Results

All experts stated that, during forensic investigations, they often have to get back to manually inspect firewall logs in textual format. So they all believe that a visual analytics system such as VAFLE would be useful and more time effective for forensic analysis. They were all willing to try it in a production environment. Their answers to the questionnaire are summarized in Tables 1 and 2.

Table 1: Expert user feedback on the usefulness of system features

Question about the usefulness of	P1	P2	P3
Heatmap Visualization	3	4	5
Line chart Visualization	3	5	5
Magic Lens Interaction	4	5	5
Activity-based clustering of IPs	4	5	5
Port-based filtering of IP activity	3	5	2

Table 2: Expert user feedback on design choices

How appropriate do you find	P1	P2	P3
Items grouping in the Magic Lens	4	4	3
Half-hour time granularity	4	3	4

7.3 Suggestions for Improvements

We also collected many suggestions towards improving VAFLE summarized as follows. Expert P1 suggested the implementation of a report generation capability in the system. P2 suggested that the system should display the topology of the network as this is an artifact network administrators often use and link it with the visualizations. P3 suggested the integration of more data sources, especially IDS alert logs. Ideally, a comprehensive forensic analysis system should provide insight into all network communication layers, whereas current industry practice involves the use of fragmented tools.

In order to enhance the anomaly detection capabilities, P1 deemed it desirable to let the user specify normal traffic for different host types and train the system to spot and flag traffic deviating from that baseline. This would involve using machine learning or, more generally, supervised classification approaches for network security. In this domain, an abundant literature can be found, including VizSec 2004,[13] although such work is not limited to network security per se. In the case of the VAST challenge datasets, such baseline traffic is unknown. We hope that future collaborations with network security practitioners with real network traffic data will allow us to explore more data mining options.

Concerning item order within the magic lens, P3 suggests implementing various orders, such as the activity levels of certain ports flagged by an IDS system. He minimized the usefulness of the port×time heatmap in the context of large organizations arguing that they usually have a strict security policy where only a few ports are left open. On the one hand, it may be useful to check that the firewall does what it is supposed to do under all circumstances. On the other hand, we believe that the analysis of all attempted connections, allowed and denied, helps understand the hacker's strategy and can inform the forensic analysis process. A typical attack would often start by fingerprinting the target infrastructure by probing it for certain vulnerabilities or characteristics such as operating system type and version, server application type and version etc. Hence, a larger number of ports that need to be included in traffic analysis than the short list of allowed services.

While discussing the types of attacks that VAFLE could potentially help identify, P3 said that DDoS attacks could only be fully identified after looking at the application logs. Therefore, VAFLE can only be used as one of several complementary tools for tracing the footsteps of a DDoS attack. P3 also mentioned data ex-filtration as one of the popular attacks that could obviously be traced with the heatmap and the time series line chart.

8. DISCUSSION

Both the use case and the user study presented in this paper lead to the conclusion that VAFLE is effective and useful for anomaly detection in firewall log events. However, we see several limitations that require further improvements.

Firstly, concerning clustering, VAFLE proposes two algorithms: k-means and spectral clustering ; the latter produces better results with non-convex clusters but suffers from a higher time complexity. Currently, clustering allows VAFLE to visually scale to large IP address spaces. This is even more important since handling IPv6 will make the visualization of much larger IP spaces more challenging. In this regard, VAFLE can be improved in two directions. On the one hand, analysts may be better served with multi-level clustering techniques that would

readily provide deep cluster hierarchies. Faster, albeit approximated, heuristics for spectral clustering[37,38] can make a good compromise between speed and cluster quality. On the other hand, better interaction techniques with clustered views could make the system more enjoyable and more usable in terms of navigation. Such techniques could include a coordinated visualization of the cluster hierarchy itself, in complement to the heatmap view in particular, in the form of an interactive dendrogram[39] or an icicle plot.[40] This would allow a more intuitive folding and unfolding of clusters as well as many editing operations through direct manipulation. The scalability of heatmaps from an interactive visualization standpoint has been achieved in other research work using optimized in-memory and on-disk data organizations.[22,41] VAFLE currently relies on clustering and the use of a modern RDBMS to further push the limits of the data it can handle.

Secondly, the user study presented in this paper surveyed domain experts opinion about the usefulness of VAFLE features. In this respect, we collected valuable feedback that confirmed many of our intuitions at the time this system was designed. However, questions related to the visual design of VAFLE remain open and require another type of evaluation to better inform design choices. With respect to the Magic Lens, we believe that including more diverse data sources could be of interest to the analyst. Other visual metaphors, e.g. pixel maps, similar to the work of Kintzel et al.[9] could be of interest here. If we keep the current stacked histogram metaphor, we may consider embedding "fisheye menu"[42] interaction inside the magic lens when too many bars must be displayed, this may often be the case in port heatmaps as in Figure 4d. Moreover, it is not clear to us whether the use of redundant data in the heatmap and line chart views is necessary or desirable, despite that domain experts deemed it useful. System robustness facing attacks specifically designed to make a certain type of visualization unusable may justify such redundancy.[43] More brainstorming with domain experts in a formative evaluation setting is required to collect more ideas. More usability assessment is required, potentially through expert reviews (in the HCI sense), to answer such questions. Guidance will be sought in recent work on the evaluation of visualizations.[34–36]

9. CONCLUSIONS

Applying visual analytics techniques to network security data helps network analysts make well-informed decisions in an accurate and time effective manner. This paper presents VAFLE, an integrated visual analytics system for firewall log events including simple yet effective visualization tools. The main contributions of this work consist in coordinated multi-level heatmap and linechart visualizations of host/port activity over time, the use of magic lens interaction to get relevant network traffic details on demand, and a discussion of custom clustering parameters that help separate rare activity patterns from dominant ones, thus guiding the analyst to a manageable subset of data to inspect. We have been successful at discovering numerous attack patterns by mining realistic datasets. A user study has established the usefulness of VAFLE from a domain expert point of view. Future work perspectives include deploying VAFLE at RESTENA, the company managing the national .lu TLD and multiple publicly and privately owned LANs country wide. Preliminary discussions and demo at their premises have confirmed the perceived usefulness of VAFLE and allowed us to identify multiple use cases in which it could be applied, beyond the firewall log analysis case described in this paper. Interestingly, IDS technology was deemed unreliable and, hence, switched off altogether by the resident network security analysts who use other network security tools to identify and mitigate threats.In the future, we plan to support multiple data sources in order to provide full end-to-end network forensics capabilities. We also plan to support live data streams. This requires revisiting existing clustering algorithms, developing data management strategies including data simplification and oblivion mechanisms as well as stable visual representations suitable for streaming data.

ACKNOWLEDGMENTS

The authors wish to thank Dr. Mahmoud Allam from Nile University for helping arrange interviews with network security experts, the latter for their time and participation in the user study as well as RESTENA network security analysts for useful discussions.

REFERENCES

[1] Marty, R., [*Applied Security Visualization*], Addison-Wesley Professional, 1 ed. (2008).

[2] Shiravi, H., Shiravi, A., and Ghorbani, A. A., "A survey of visualization systems for network security," *IEEE Transactions on Visualization and Computer Graphics* **18**, 1313–1329 (Aug. 2012).

[3] Conti, G., Abdullah, K., Grizzard, J., Stasko, J., Copeland, J., Ahamad, M., Owen, H. L., and Lee, C., "Countering security information overload through alert and packet visualization," *Computer Graphics and Applications, IEEE* **26**(2), 60–70 (2006).

[4] Goodall, J. R., Lutters, W. G., Rheingans, P., and Komlodi, A., "Preserving the big picture: Visual network traffic analysis with tnv," in [*Visualization for Computer Security, 2005. (VizSEC 05). IEEE Workshop on*], 6 (2005).

[5] Ball, R., Fink, G. A., and North, C., "Home-centric visualization of network traffic for security administration," in [*Proceedings of the 2004 ACM workshop on Visualization and data mining for computer security*], *VizSEC/DMSEC '04*, 55–64, ACM, New York, NY, USA (2004).

[6] Fischer, F., Mansmann, F., Keim, D. A., Pietzko, S., and Waldvogel, M., "Large-scale network monitoring for visual analysis of attacks," in [*Proceedings of the 5th international workshop on Visualization for Computer Security*], *VizSec '08*, 111–118, Springer-Verlag, Berlin, Heidelberg (2008).

[7] Lakkaraju, K., Yurcik, W., and Lee, A. J., "Nvisionip: netflow visualizations of system state for security situational awareness," in [*Proceedings of the 2004 ACM workshop on Visualization and data mining for computer security*], *VizSEC/DMSEC '04*, 65–72, ACM, New York, NY, USA (2004).

[8] Mansmann, F., Keim, D., North, S., Rexroad, B., and Sheleheda, D., "Visual analysis of network traffic for resource planning, interactive monitoring, and interpretation of security threats," *Visualization and Computer Graphics, IEEE Transactions on* **13**(6), 1105–1112 (2007).

[9] Kintzel, C., Fuchs, J., and Mansmann, F., "Monitoring large ip spaces with clockview," in [*Proceedings of the 8th International Symposium on Visualization for Cyber Security*], *VizSec '11*, 2:1–2:10, ACM, New York, NY, USA (2011).

[10] Lamagna, W. M., "An integrated visualization on network events vast 2011 mini challenge #2 award: "outstanding integrated overview display"," in [*Visual Analytics Science and Technology (VAST), 2011 IEEE Conference on*], 319–321 (2011).

[11] McPherson, J., Ma, K.-L., Krystosk, P., Bartoletti, T., and Christensen, M., "Portvis: a tool for port-based detection of security events," in [*Proceedings of the 2004 ACM workshop on Visualization and data mining for computer security*], *VizSEC/DMSEC '04*, 73–81, ACM, New York, NY, USA (2004).

[12] Boschetti, A., Salgarelli, L., Muelder, C., and Ma, K.-L., "Tvi: a visual querying system for network monitoring and anomaly detection," in [*Proceedings of the 8th International Symposium on Visualization for Cyber Security*], *VizSec '11*, 1:1–1:10, ACM, New York, NY, USA (2011).

[13] [*VizSEC/DMSEC '04: Proceedings of the 2004 ACM workshop on Visualization and data mining for computer security*], ACM, New York, NY, USA (2004). 100046.

[14] Shneiderman, B., "The eyes have it: A task by data type taxonomy for information visualizations," in [*Proceedings of the 1996 IEEE Symposium on Visual Languages*], *VL '96*, 336–, IEEE Computer Society, Washington, DC, USA (1996).

[15] Conti, G., [*Security Data Visualization*], No Starch Press, San Francisco, CA, USA (2007).

[16] Fekete, J.-D., "The infovis toolkit," in [*Proceedings of the IEEE Symposium on Information Visualization*], *INFOVIS '04*, 167–174, IEEE Computer Society, Washington, DC, USA (2004).

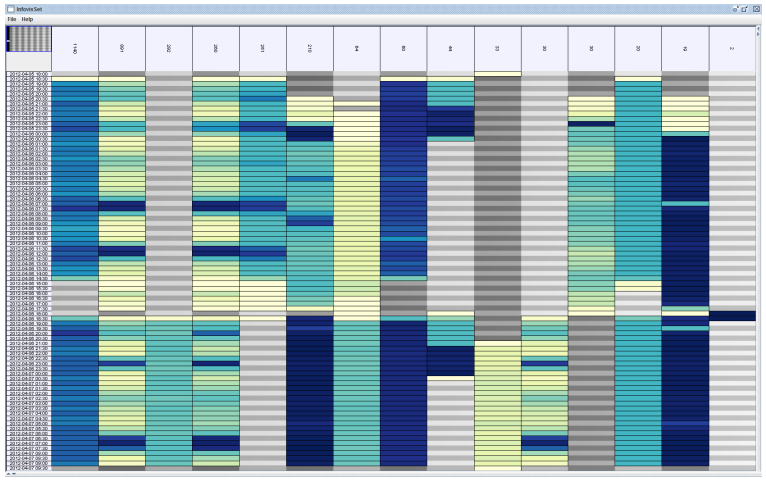
[17] Furnas, G. W., "Generalized fisheye views," in [*Proceedings of the SIGCHI conference on Human factors in computing systems*], *CHI '86*, 16–23, ACM, New York, NY, USA (1986).

[18] Fekete, J.-D. and Plaisant, C., "Excentric labeling: dynamic neighborhood labeling for data visualization," in [*Proceedings of the SIGCHI conference on Human factors in computing systems: the CHI is the limit*], *CHI '99*, 512–519, ACM, New York, NY, USA (1999).

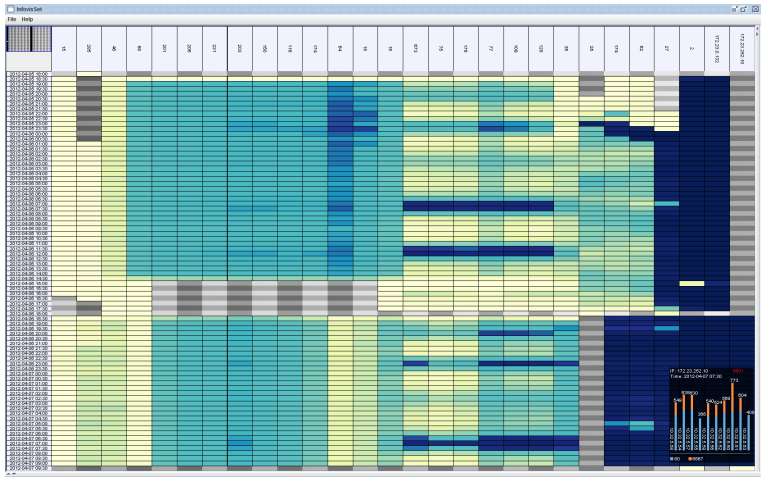
[19] Ahlberg, C., Williamson, C., and Shneiderman, B., "Dynamic queries for information exploration: an implementation and evaluation," in [*Proceedings of the SIGCHI conference on Human factors in computing systems*], *CHI '92*, 619–626, ACM, New York, NY, USA (1992).

[20] Wilkinson, L. and Friendly, M., "The History of the Cluster Heat Map," *The American Statistician* **63**, 179–184 (May 2009).

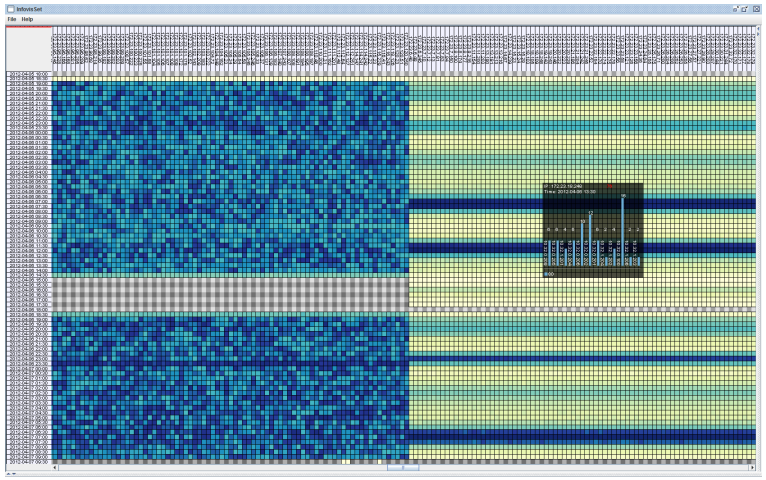
[21] Ghoniem, M., Cambazard, H., Fekete, J.-D., and Jussien, N., "Peeking in solver strategies using explanations visualization of dynamic graphs for constraint programming," in [*Proceedings of the 2005 ACM symposium on Software visualization*], *SoftVis '05*, 27–36, ACM, New York, NY, USA (2005).
[22] Van Ham, F., "Using multilevel call matrices in large software projects," in [*Proceedings of the Ninth annual IEEE conference on Information visualization*], *INFOVIS'03*, 227–232, IEEE Computer Society, Washington, DC, USA (2003).
[23] Eisen, M. B., Spellman, P. T., Brown, P. O., and Botstein, D., "Cluster analysis and display of genome-wide expression patterns," *Proceedings of the National Academy of Sciences (PNAS)* **95**, 14863–14868 (December 1998).
[24] Wu, H.-M., Tien, Y.-J., and Chen, C.-h., "Gap: A graphical environment for matrix visualization and cluster analysis," *Comput. Stat. Data Anal.* **54**, 767–778 (Mar. 2010).
[25] Bertin, J., [*Semiology of graphics*], University of Wisconsin Press, Madison, Wis. (1983).
[26] Liiv, I., "Seriation and matrix reordering methods: An historical overview," *Stat. Anal. Data Min.* **3**, 70–91 (Apr. 2010).
[27] Mohar, B., "The laplacian spectrum of graphs," in [*Graph Theory, Combinatorics, and Applications*], 871–898, Wiley (1991).
[28] Mohar, B., "Laplace eigenvalues of graphs – a survey," *Discrete Mathematics* **109**(1–3), 171 – 183 (1992).
[29] Mardia, K. V., Kent, J. T., and Bibby, J. M., [*Multivariate Analysis*], Academic Press (1979).
[30] Tan, P.-N., Steinbach, M., and Kumar, V., [*Introduction to Data Mining, (First Edition)*], Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA (2005).
[31] Lee, H., Kihm, J., Choo, J., Stasko, J., and Park, H., "ivisclustering: An interactive visual document clustering via topic modeling," *Computer Graphics Forum* **31**(3pt3), 1155–1164 (2012).
[32] Bier, E. A., Stone, M. C., Pier, K., Fishkin, K., Baudel, T., Conway, M., Buxton, W., and DeRose, T., "Toolglass and magic lenses: the see-through interface," in [*Conference companion on Human factors in computing systems*], *CHI '94*, 445–446, ACM, NY, USA (1994).
[33] Kosara, R., Healey, C. G., Interrante, V., Laidlaw, D. H., and Ware, C., "Thoughts on User Studies: Why, How, and When," *Computer Graphics and Applications* **23**, 20–25 (July 2003).
[34] Lam, H., Bertini, E., Isenberg, P., Plaisant, C., and Carpendale, S., "Empirical studies in information visualization: Seven scenarios," *Visualization and Computer Graphics, IEEE Transactions on* **18**(9), 1520–1536 (2012).
[35] Sedlmair, M., Meyer, M., and Munzner, T., "Design study methodology: Reflections from the trenches and the stacks," *IEEE Transactions on Visualization and Computer Graphics* **18**(12), 2431–2440 (2012).
[36] Tory, M. and Moller, T., "Evaluating visualizations: do expert reviews work?," *Computer Graphics and Applications, IEEE* **25**(5), 8–11 (2005).
[37] Yan, D., Huang, L., and Jordan, M. I., "Fast approximate spectral clustering," in [*Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*], *KDD '09*, 907–916, ACM, New York, NY, USA (2009).
[38] Sakai, T. and Imiya, A., "Fast spectral clustering with random projection and sampling," in [*Machine Learning and Data Mining in Pattern Recognition*], Perner, P., ed., *Lecture Notes in Computer Science* **5632**, 372–384, Springer Berlin Heidelberg (2009).
[39] Seo, J. and Shneiderman, B., "Interactively exploring hierarchical clustering results," *Computer* **35**(7), 80–86 (2002).
[40] Ghoniem, M. and Fekete, J.-D., "Matrix view of graphs and direct manipulation of cluster hierarchies," in [*Proceedings of the 15th French-speaking conference on human-computer interaction*], *IHM 2003*, 206–207, ACM, New York, NY, USA (2003).
[41] Bohn, S. J., Payne, D., Nakamura, G., and Love, D., "Analytics for massive heat maps," in [*Proceedings of Visualization and Data Analysis 2009*],
[42] Bederson, B. B., "Fisheye menus," in [*Proceedings of the 13th annual ACM symposium on User interface software and technology*], *UIST '00*, 217–225, ACM, New York, NY, USA (2000).
[43] Conti, G., Ahamad, M., and Stasko, J., "Attacking information visualization system usability overloading and deceiving the human," in [*Proceedings of the 2005 symposium on Usable privacy and security*], *SOUPS '05*, 89–100, ACM, New York, NY, USA (2005).



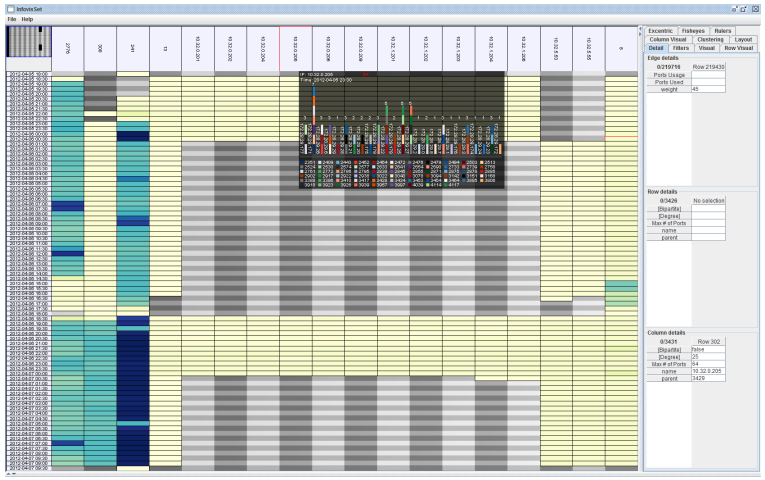
(a) Heatmap of hosts activity over time: Clustered view all clusters folded



(b) Clustered heatmap showing the top talkers at the right, with outliers expanded

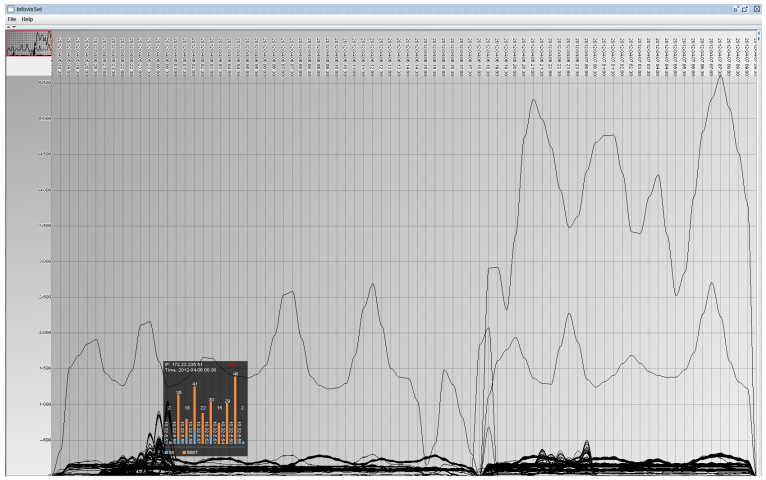


(c) Heatmap of hosts activity over time: Close-up view of individual host activity at the junction of the two left-most expanded clusters.

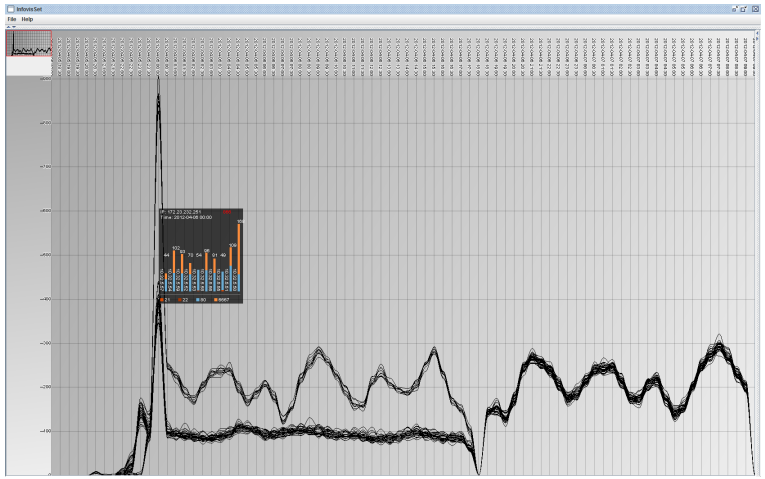


(d) Portscan attack characterized by a colorful magic lens. The side panel provides extra textual information about selection.

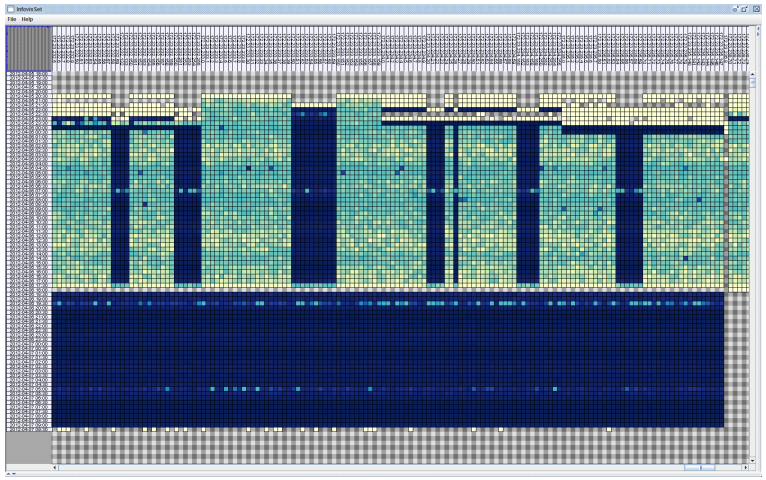
Figure 3



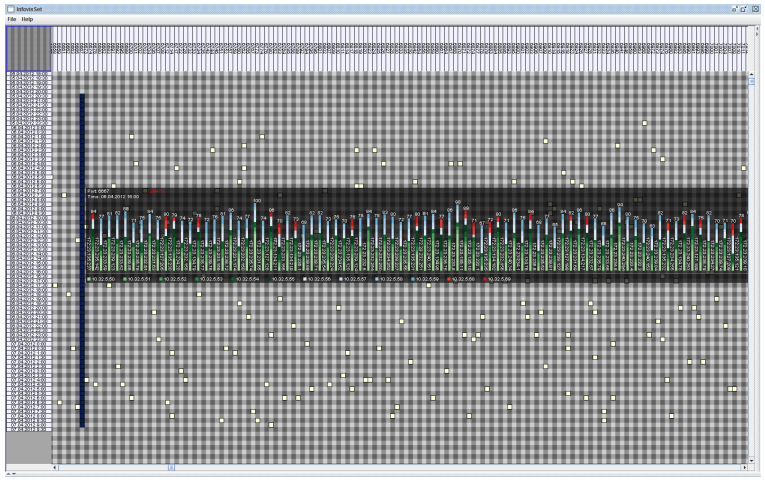
(a) Time series view showing top talkers (upper half) and botnet activity (bottom) with detailed port activity through the magic lens.



(b) Botnet activity orchestrated through IRC



(c) Spread of suspicious web traffic over time (in dark blue) from day 1 to day 2. Potential data ex-filtration attack.



(d) Ports heatmap detailing IRC traffic

Figure 4