IMap: Visualizing Network Activity over Internet Maps

J. Joseph Fowler
Michael Schneider
Thienne Johnson
Carlos Acedo
Loukas Lazos
Paolo Simonetto*
Stephen Kobourov

{fowler,thienne,paolosimonetto,mschneid,cajac2,kobourov,llazos}@email.arizona.edu
University of Arizona, Tucson, Arizona, US, 85721

ABSTRACT

We propose a novel visualization, IMap, which enables the detection of security threats by visualizing a large volume of dynamic network data. In IMap, the Internet topology at the Autonomous System (AS) level is represented by a *canonical map* (which resembles a geographic map of the world), and aggregated IP traffic activity is superimposed in the form of *heat maps* (intensity overlays). Specifically, IMap groups ASes as contiguous regions based on AS attributes (geo-location, type, rank, IP prefix space) and AS relationships. The area, boundary, and relative positions of these regions in the map do not reflect actual world geography, but are determined by the characteristics of the Internet's AS topology. To demonstrate the effectiveness of IMap, we showcase two case studies, a simulated DDoS attack and a real-world worm propagation attack.

Categories and Subject Descriptors

C.2.0 [**Computer-Communication Networks**]: General | Security and protection; C.3.8 [**Computer Graphics**]: Application

General Terms

Security

Keywords

topology visualization, network, anomaly, security, map

1. INTRODUCTION

Network managers face the challenging task of continuously monitoring their networks for suspicious activities. The volume of network data to be inspected can be enormous, especially when performing inspection at the packet-level. The data often originates from disparate sources and is noisy when malicious and normal traffic co-occur. A well-integrated and well-designed visualization system greatly facilitates the effective presentation and interactive analysis of the underlying high-volume and complex data. However, most visualization tools in this setting show a great deal of rapidly changing data, resulting in visual representations matching the complexity of the original data [2, 16]. Consequently, security experts tend to rely on easy-to-understand graphics such as line plots, box-plots, and pie-charts.

To cope with the visual complexity of large datasets, we developed IMap, a simple and intuitive network visualization. It relies on the map metaphor, which has been successfully used to visualize relational datasets [11] and naturally lends itself to depicting network activity and attributing/correlating security threats to their origins using a familiar visual paradigm. IMap visualizes the "security posture" of a monitored network without overwhelming the cognitive ability of human analysts; see Fig. 1.

This is achieved by superimposing *heat maps* (colorized intensity overlays) of network activity onto synthetic geographic-like *canonical maps*. These maps represent the Internet topology at the Autonomous System (AS) level. An AS

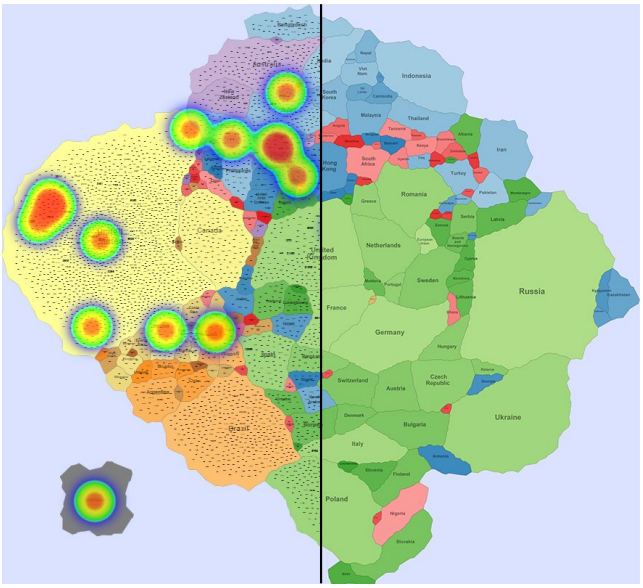


Figure 1: Map of the AS topology with and without nodes and hot spots overlaid. Each country is shown as a contiguous region containing the ASes operating in their territory.

*The first three authors contributed to the paper equally.

Vizsec '14 Paris, France

is a group of computer networks, typically under the same administrative authority, using the same routing policy [23]. Business relationships on the AS level govern the flow of network traffic across the Internet. The Internet topology can be abstracted to an AS topology graph [3]. The visualization of the Internet's AS topology is important to security analysts for many reasons, such as understanding peering relationships, routing hot spot detection, early evaluation of the attack severity, localizing the source of an attack, understanding attack propagation patterns, and others [3, 21, 24].

Representing the AS topology as a synthetic geographic map, as opposed to overlaying data onto a physical geographic map, offers significant advantages. The latter are dominated by geographies lacking interest, *e.g.*, oceans and countries with limited Internet presence. Moreover, the surface area of each country can be minimally related to the country's presence and contribution towards global traffic. The physical distance between two countries in a geographic map does not necessarily correspond to the level of connectivity between the countries. On the other hand, in IMap, country size is proportional to the importance of the respective ASes in the AS topology. Therefore, countries of small geographic area but significant role on the AS graph, occupy larger area in the IMap (*e.g.*, Ukraine) and *vice versa* (*e.g.*, Greenland); see Fig. 1. Moreover, the distance between two countries in IMap is related to the level of connectivity between the ASes in the corresponding geographic countries.

Contributions

We represent the Internet AS topology as a geographic-like map, but one in which country sizes and relative placement are based on the structural properties of the Internet. We employ an Intrusion Detection System (IDS) that computes anomaly scores based on live IP traffic streams, aggregated at the AS level, and visualize the IDS anomaly scores as heat map overlays. We demonstrate the effectiveness of IMap with two case studies: a simulated DDoS attack and a real-world worm propagation attack. A companion website[1] contains high resolution images of the figures in this paper, as well as a video of the network analysis described in Section 5.

2. RELATED WORK

Since our work spans several fields of network analysis and information visualization, we cannot survey all related work. We highlight several approaches and systems that most influenced our work.

AS topology is traditionally visualized with static node-link diagrams. The AS Level Internet Graph [7] depicts the AS topology in polar coordinates, using the out-degree of an AS to determine the distance from the center of a circle, and its geographic location to determine its position around the circle. Cyclops [22] shows the topology as a graph and allows the user to focus on different areas. Node size is proportional to connectivity, to differentiate big ISPs from small ones, and edge thickness is proportional to the age of a link. Using complexity reduction of the AS topology dataset results in smaller graphs and better screen space utilization, where instead less essential aspects of the graph are shown with different modes of representation. In VAST [21], a quad-tree based visualization represents the AS space in the plane, where the third dimension shows the value of several AS metrics. As in these previous approaches, we also use nodes and links to display ASes and the connections between them. However, we also show the grouping of ASes into countries and capture the importance and high-level structure with the size of the regions and their relative placement.

Much work has been devoted to creating clear and informative graph drawings by optimizing node positions [10]. Studies have shown that visual embellishments, applied to charts and other visualizations, may memorization and recall [1], though hinder speed of visual search in some cases [4]. Space-driven partitioning using a grid can have better short-term performance in the revisitation of graph nodes than a detail-driven partitioning using Voronoi diagrams [12]. Clustered data can be represented by colored point clouds (node-diagrams), colored network (node-link-diagram), or with a landscape metaphor (node-link-group diagram). Although the most complex of the three, and perhaps because of their familiarity, landscape representations [9, 11, 26] have been shown to perform well [28]. Fragmented regions may cause misinterpretation [14], but if each group is represented by a single, contiguous region, these representations provide only benefits over node-link diagrams [25]. Hand-drawn maps with contiguous regions have already been used to depict different aspects of the web,[2] but creating such maps relies on artistic talent, knowledge and wit. For use in network security visualization it is necessary to automate the production of such maps with an eye on aesthetics, but some of the genius of the hand-made maps will be missing.

Basic network activity visualization systems also employ techniques such as treemaps [2] and rings [30]. Parallel coordinate plots can show individual dimensions or fields of the dataset such as TCP source port, source IP address, destination IP address, and TCP destination port [13]. Some of these approaches can result in poor resolution, when displaying quantities (such as packets sent), and difficulty in analyzing the visualization, as the number of connections increases. Some comprehensive network activity tools combine several basic data views. A visual pattern detection tool, showing temporal activity for thousands of hosts at once, builds upon the structural properties of IP addresses belonging to subnets and a global prefix, therefore describing a 2-level hierarchy [16]. Every visual item (host) shows temporal activity (traffic) as a small 24-hour clock, but for large networks, visualizing each host becomes difficult. Anomalous AS activity is detected with an interactive exploration of a choropleth map of the world, where the saturation reflects the average malicious score of each AS in the country, by the appropriate use of graphical aspects (*e.g.*, size, position, shape, color) and network features (*e.g.*, number of malicious servers, geographical location, AS size, types of malicious activities) [24].

Contrary to most approaches, in IMap, flows are processed per AS, thus creating fewer flows for analysis. This reduces the clutter in the user interface due to the fewer data points to be displayed as hot spots over the canonical map. As a result, the cognitive ability of human analysts is not overwhelmed. Differently from [24], which also visualizes the AS topology over a geographic map, IMap uses a synthetic geography that accurately represents the underlying AS graph structure. This conveys visual information on the role of each AS or a global scale.

[1] `http://vizsec-gama.cs.arizona.edu/`

[2] `http://xkcd.com/195/`, `http://xkcd.com/256/`

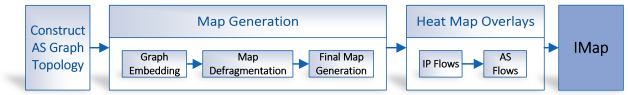


Figure 2: IMap overview

3. DIAGRAMS GENERATION

The IMap diagrams are built in a three-stage process shown in Fig. 2. The first stage deals with the construction of the AS topology graph from the given AS data [8]. The second stage is responsible for embedding the graph in the plane and for creating the canonical AS map. The final stage overlays network activity as heat maps.

3.1 Construction of the AS Graph Topology

IDSes capture IP flows at the border of monitored networks. These traces of network activity contain a wealth of information used in attack detection, *e.g.*, source and destination addresses and ports, the packet types exchanged between hosts, *etc.* However, the size of the IP space is prohibitively large for producing any meaningful trace visualization. Moreover, the IP space does not directly reflect the Internet topology. We exploit the organization of the IP space into ASes to significantly reduce the number of visualized data points by visualizing the AS graph.

Even at the AS level, the number of data points related to network activity is large: in July 2014, the number of registered ASes was 66 710 [20]. We classify ASes as customer ASes (stub or multi-homed) and ISPs (transit AS) to further reduce the data. Stub ASes are end-networks that originate and receive traffic, but do not relay traffic. On the other hand, transit ASes relay traffic from stub ASes. The majority of ASes fall into the former category, while transit ASes form the core of the Internet. Aggregation of stub and multi-homed customer ASes to their parent ASes (ISPs) results in a highly-connected network mesh with fewer nodes.

In the first step we extract the AS graph $G(V, E)$, where the vertex set V represents the set of ASes and the edge set E represents AS peering relationships. Both V and E can be inferred from BGP path vector advertisements [19]. A BGP path vector is of the form $\mathrm{AS}_0, \mathrm{AS}_1, \ldots, \mathrm{AS}_k$. An adjacency between AS_i and AS_j on a BGP path vector indicates a peering relationship between the respective ASes. Note that although AS_i can have a peering relationship with AS_j, this does not necessarily imply the existence of a direct physical link between these ASes (as the two might be connected through a multi-hop route).

Several additional AS attributes can be extracted or computed from BGP path advertisements. These include the AS type (stub, multi-homed, transit, T1, large/small ISP) and the IP prefixes advertised by each AS. For more details about the extraction methods of AS topology attributes, see [19]. We use these additional attributes to assign weights to both the vertices and edges of G. These weights are meant to represent the relative importance of each AS in the Internet hierarchy. Specifically, we compute weights using the concept of an AS customer cone [19]. The customer cone of AS_i is defined as the set of ASes that AS_i can reach using customer links (excluding all p2p links). For node weights, we use the *AS cone-size weight*, which is the number of ASes belonging to the customer cone of the node AS_i. Intuitively, the larger the number of ASes served by AS_i, the higher the importance of that AS in the Internet topology, which results in a higher weight being assigned to AS_i.

3.2 Generation of the Canonical Map

During the generation of the canonical map that represents the AS graph G we ensure that (1) the ASes of the same geographical country are placed within the same country in the canonical map, (2) each country in the canonical map is contiguous, with area proportional to the number of ASes located in that country, and (3) countries that are closely connected on the AS level correspond to countries that are close to each other[3] in the canonical map.

3.2.1 Graph Embedding

The AS graph G is drawn on the plane using the multiscale spring embedder `sfdp` of the GraphViz suite,[4] which tends to place highly-interconnected ASes within a natural cluster, *e.g.*, a geographical country or continent, in close proximity to one other. The `sfdp` algorithm has been chosen due to the large size of the input graphs (about 10k nodes and 100k edges), which requires an efficient, scalable embedder. Using different edge weights in the graph affects how well a spring embedder groups nodes of the same country. When an emphasis is placed on edges with heavier weights, we obtain a drawing where the heavier weighted ASes in close proximity are co-located in the center of the map, irrespective of the country origin of an AS. The remaining nodes are placed in less meaningful positions in the map, as they have less of an impact on the drawing. When (in the next step) ASes of the same geographic country are grouped into one contiguous country in the canonical map, not only is the proximity between the heaviest ASes distorted, but there is also a side effect of creating an artificial proximity between less important ones. On the other hand, when all edges have equal weight, we obtain an embedding that emphasizes local AS relationships. We determined (by analyzing cluster modularity) that using equally-weighted edges yields the best cluster-respecting results; see Fig. 3. Note that this choice penalizes connections between larger ASes, which carry a considerably larger traffic volume.

3.2.2 Map Defragmentation

The graph embedding obtained in the previous step is likely to place nodes of the same geographic country relatively far apart of each other, in order to satisfy the embedding criteria. This can lead to *fragmented* countries, *i.e.*, separated into multiple disjoint regions. We obtain a contiguous map using the cluster-based approach (CBA) in [17].

In CBA, nodes of the same country are initially attracted towards the barycenter of that country, *i.e.*, the average position of the country nodes. The nodes are surrounded by an uncrossable but flexible boundary, and then moved towards their original positions using a force-directed algorithm. As a result, nodes of the same country are placed in a contiguous region of the plane, while preserving the initial node distances and relative positions as much as possible.

Compare the fragmented map in Fig. 3 to its defragmented counterpart in Fig. 1. Observe that the fragmentation in Fig. 3 is extreme. Given the consistent continent-level col-

[3]W. Tobler, 1st law of geography: "Everything is related to everything else, but near things are more related than distant things".

[4]Available for free download from `www.graphviz.org`

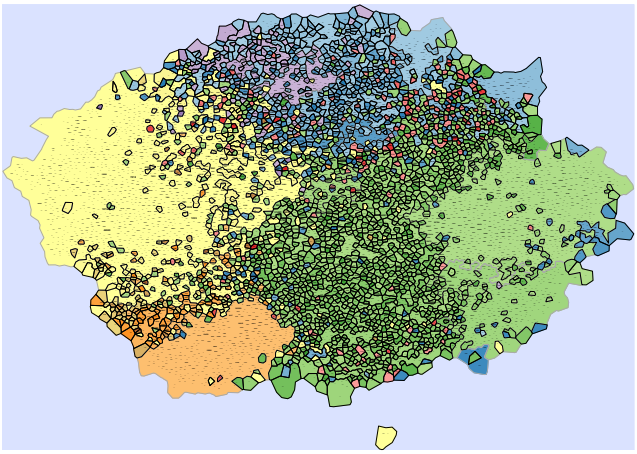


Figure 3: Topology map generated without applying the defragmentation step of Section 3.2.2.

oring scheme in Table 1, we can visually verify that local AS connections shown in Fig. 3 are mostly preserved in Fig. 1, as most of the countries occupy the same region of the drawing. This indicates that the defragmentation process preserves the country proximity produced by the graph embedder.

3.2.3 *Final Diagram Generation*

In the final stage, the embedded, clustered graph is converted into a map using the GMap framework [11], as implemented in `gvmap` from the GraphViz suite. This consists of creating a Voronoi diagram of the graph nodes and merging neighboring Voronoi cells that belong to the same cluster into contiguous regions.

During this stage, the AS labels are scaled to depict the AS weights. Given the large variance in node weights, a logarithmic scaling is used for font sizes. This scaling facilitates the identification of more important ASes. We also add the most relevant AS graph edges and assign them an alpha channel value that depends on their weight: high-weight edges are visible, while low-weight edges are transparent. Finally, we label each country and we insert an *Unknown* country in an unused portion of the drawing. The *Unknown* country is used to display activity coming from IP addresses that cannot be mapped to any ASes, which can occur for several reasons (e.g., due to IP spoofing attacks).

The countries are colored according to their continent. Each continent is associated with a range between two colors, extracted from a ColorBrewer[5] qualitative pastel palette. Countries are then associated with a color in that range according to an ordering based on population. The range of colors associated to each continent are reported in Table 1.

[5]`http://www.colorbrewer.org`

Continent	First Color	Last Color
North America	Light Yellow	Brown
South America	Light Orange	Dark Orange
Europe	Light Green	Dark Green
Asia	Light Blue	Dark Blue
Oceania	Light Purple	Dark Purple
Africa	Light Red	Dark Red

Table 1: Colors associated with each continent.

3.3 Heat Map Overlays

Heat maps are a well-known tool for visualizing dynamic data over a geographical map, *e.g.*, temperature variation, wave height, wind speed, *etc.* In IMap, heat maps are used to display overlays of any feed from an IDS aggregated at the AS level, *e.g.*, network traffic statistics, network events, utilization of resources, *etc.* To aggregate information at the AS level, IDS warnings should map source IP nodes to their respective AS. We designed an IDS system that aggregates IP traffic per AS and performs per-AS flow analysis to detect anomalies [15]. The proposed anomaly detection system employed the distance metrics methodology [29]. Information distance (or divergence) is a measure of the difference between probability distributions, which model network traffic attributes of interest. We adopted a semi-supervised statistical anomaly detection technique.

In IMap, heat maps are generated using the javascript libraries Heatmap.js[6] and OpenLayers.[7] Each hot spot is represented by a color gradient. The maximum intensity is represented by the color red (hot) , and the minimum is represented by the color blue (cold). The color gradient is designed so that a broader range of colors is assigned to higher intensity values, thus avoiding emphasis on unimportant information.

4. MAP ANALYSIS

In this section, we report and analyze several statistics about the generated maps.

Country Area

Table 2 lists the correlation coefficients between country statistics and country area in the map. We considered two values for country area in the drawing: (i) the area of the polygon associated with each country and (ii) the area of its convex hull. A convex hull is the smallest convex polygon that contains the original one, and therefore, always has greater than or equal area compared to the original polygon.

Note that the correlations for the polygon-area and convex-hull-area are nearly identical; see Table 2. The areas of the convex hulls are only 7.6% larger on average indicating that the polygons are mostly convex. Convex contours are generally believed to be easier to identify in a drawing, to the point of being sometimes enforced in diagrams that share the same foundations with IMap [27].

Also observe that the polygon area is strictly correlated to the number of nodes in the topology graph as well as to the sum of the weights of the AS nodes in the graph (second

[6]`http://www.patrick-wied.at/static/heatmapjs/`
[7]`http://openlayers.org/`

	Polygon Area	Convex Hull Area
Total ASes	0.9256	0.9364
Number AS nodes	0.9869	0.9895
Total node weights	0.9849	0.9879
Country area	0.6729	0.6625

Table 2: Correlation between country statistics and their areas in the generated IMap.

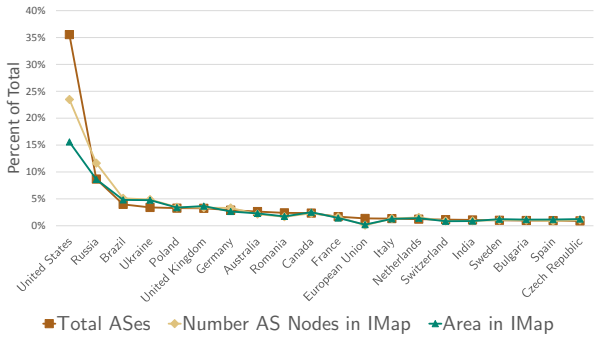


Figure 4: Statistics for top 20 countries with the most ASes.

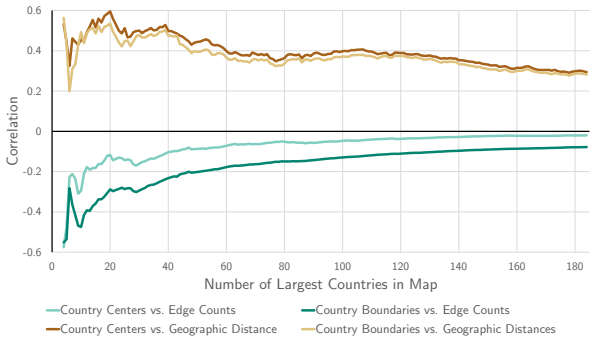


Figure 5: Distance-based correlations restricted to the k largest countries in the map, where $k \in \{4, 5, \ldots, 184\}$.

and third lines in Table 2). This indicates that the area of the countries can be reliably used to estimate these statistics directly on the map. The correlation decreases, though still remains fairly high (*i.e.*, $> .9$) when considering the total number of ASes for a country (first line in Table 2). This includes ASes not present in the AS topology graph. Not all ASes may be actively advertising BPG paths that we can deduce. Figure 4 visually illustrates the strength of these correlations for the total number of ASes, the number of ASes is the graph, and the area of each country for the top 20 countries with the most ASes each. The nearly overlapping lines indicate high correlation. Furthermore, they signify that the area of the IMap is strongly correlated to both the total ASes as well as the number of ASes in the IMap.

Country Distances and Adjacencies

Figure 5 provides correlations between country distances on the map, the physical geographic distances, and the number of AS topology graph edges that connect them. The correlation is computed considering each pair of countries among the k countries with the largest areas in the map, where k ranges from 4 (United States, Russia, Brazil, and Ukraine) to 184 (all countries in the map).

We evaluate two types of country distances: between *country centers* (centroids of the country polygons), and between *country boundaries* (minimal distance between any two vertices of the country polygons).

As expected, the correlation between number of connecting edges and map distance is negative, which indicates that highly-interconnected countries are in close proximity. These characteristics are well respected when considering the largest countries in the map (highest inverse correlation of 0.55), but it is virtually absent when considering all the countries in the drawing. This may be due to the lack of direct AS links between most pairs of countries (given that relatively few countries

we would need to place such pairs of countries infinitely away from each other, which is not feasible in practice. Instead, IMap optimizes the drawing space by placing such countries as far as possible, subject to the constraint that all countries form a single contiguous continent.

Therefore, although not completely faithful, country proximity provides insights about the connectivity of the most prominent countries. With this in mind, we believe that boundary-distance is better than center-distance, for encoding country connectivity. The correlation between distances in our map and real world distances is also fairly high, indicating that the map distances partially reflect geographical ones. This suggests that the AS connectivity is significantly influenced by physical geography, at least when considering all connections as equally important. Despite not being directly enforced, this property has interesting side effects on the usability of IMap, as it likely helps locate countries and increases map memorability.

Computation Time

The computation of the maps in this paper with 6940 nodes and 63 845 edges required an hour on a PC with an Intel i7 processor. The vast majority of the time was taken by the canonical map computation, and only minimal time was required to extract data for the heat map overlays. However, since the AS topology changes slowly, updated maps are only expected to be generated infrequently (*e.g.*, once a month).

5. CASE STUDIES

Two scenarios were created to demonstrate the usefulness of IMap. First, we use a sequence of heat maps to show the evolution of a DDoS attack from the perspective of a monitored network. In the second scenario, we used real data from a worm propagation event [6] to study the origins of the worm and its propagation patterns. We used the CAIDA UCSD IPv4 Routed /24 Topology Dataset [8] to build the underlying AS topology in the IMap generation process.

5.1 DDoS Attack

We generated synthetic DDoS attacks of varying intensity against a monitored network over a period of 25 minutes, using several attack topologies. Table 3 specifies the traffic parameters for each 5-minute interval. Background traffic was generated by the D-ITG traffic generator[8] (100 random source IP nodes) and the DDoS attack (volumetric attack) was generated by the *bonesi*[9] package. The data rates from the DDoS hosts were greater than those generated by background traffic sources (in terms of the number of packets, volume, and number of IP flows).

To detect anomalous traffic relative to a monitored network, the IDS operates in two phases: a training phase and an online phase. During the training phase, we create stochastic models of normal network activity in the form of

[8] `http://traffic.comics.unina.it/software/ITG`
[9] `http://code.google.com/p/bonesi`

Interval	Event
1	Three origin AS - high rate and number of flows
2	Four ASes - high rate, number of flows and volume
3	Several ASes - Botnet attack
4	Three ASes - high rate and number of flows
5	Several ASes - Weaker Botnet attack

Table 3: Traffic parameters per 5-minutes intervals

probability distributions. We use training datasets for this purpose. During the online phase, which is shown in Fig. 6, we first create empirical probability distributions of both the incoming and outgoing traffic intercepted at the capturing point using count-based histograms. We then compare the online distributions with those obtained during the training phase and compute the Jeffrey distance [29] between the respective distributions. This distance is used to measure the deviation of live traffic patterns from normal ones for relevant traffic metrics (*i.e.*, packet count, traffic volume, number of IP flows, *etc.*). We normalize the distance so that the dynamic range of the metric is between zero and one. Finally, we combine the Jeffery distances of different traffic metrics (*e.g.*, packet count and number of IP flows) to create composite metrics of the network activity. This weighted combination allows the detection of different types of DDoS attacks [15, 18]. Specifically, composite metric C_1 combines anomaly scores related to packet count and number of IP flows. Composite metric C_2 combines anomaly scores related to traffic volume and number of IP flows (more details in [15]). Composite metrics are then compared to a threshold τ, which defines the values for coloring the heat maps.

Figure 7 shows the heat map sequence resulting from the analysis of five 5-minute intervals of network activity. All heat maps show the origin of the traffic relative to the monitored network, *i.e.*, the nodes that appear in the IMap are the sources of the attacks. For each map along the first line in Fig. 7, there is a corresponding map along the second line (metrics C_1 and C_2 in each line, resp.). All heat values with anomaly scores less than $\tau = 0.7$ were suppressed. This threshold yielded the best trade off between DDoS attack detection and suppression of false alarms.

We observe that the security analyst can evaluate the severity of the anomaly by the color of the hot spot—nodes with a green hot spot are above the threshold but not near the maximum value, while nodes with a red hot spot are at the maximum value. The analyst can tune the heat map parameters (*i.e.*, coloring scale and τ) to control the number of hot spots and coloring degrees shown on the map.

The traffic aggregation at the AS level, as opposed to the IP level, reduces the number of hot spots that appear on the map. Instead of showing potentially thousand hot spots, the security analyst has to visually process a much smaller set. For instance, around 20 ASes represent thousands of nodes during large-scale botnet attacks. Moreover, the security analyst infers that most attacking machines belong to only a few ASes, which may be an aftermath of ingress filtering at the edge gateways, which blocks outgoing traffic from invalid source IPs. The *Unknown* country (represented by the island at left lower corner) aggregates all the nodes not mapped to any AS, *i.e.*, traffic sent from spoofed IPs or unadvertised IP prefixes. In every interval, this aggregate node has a hot spot, signifying traffic of unknown origin.

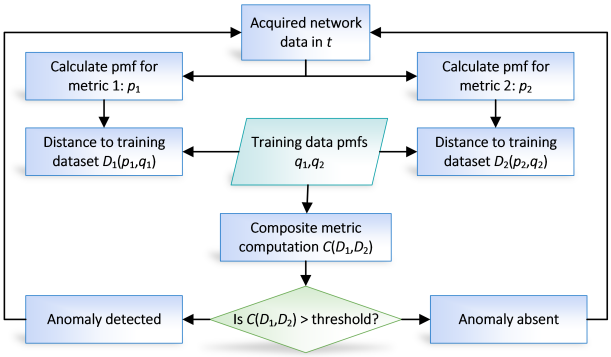


Figure 6: Flowchart of the anomaly detection process

We emphasize that our goal here is to demonstrate the visual benefits of IMap and not optimize the IDS that performs anomaly detection. IMap can operate in tandem with any IDS (or other security tool or sensing mechanism) that attributes anomaly metrics at the AS level.

5.2 Worm Propagation

In July 19th, 2001, a variant of the Code-Red worm appeared and spread very rapidly around the world. The CAIDA Code-Red Worms dataset [6] contains packet headers collected from three different network monitors. In the animations provided by CAIDA [5], the worm spread is presented by heat maps overlaid on top of geographical maps. Their conclusion was that "physical and geographical boundaries are meaningless in the face of a virulent attack". We used one of the datasets containing the data relative to the nodes (IP addresses and their respective Autonomous System) that were observed to be transmitting the worm.

The dataset was processed so each hot spot represents one Autonomous System with reported activity in a given time interval. For this case study, a positive identification of a node participating in the spread (data pre-computed in the CAIDA dataset) received a heat value equal to 1, so it would create a hot spot with maximum intensity. We generated several sequences of heat maps, representing different time intervals (one second, one minute,[10] and one hour). The interval size provides different perspectives on the analysis of the dataset. Figure 8 shows the spread of the worm at various times during the propagation.

When visually inspecting the map, the size of the labels aid in identifying the type of the AS presenting the anomaly, in terms of importance in the network (node weights). The security analyst can visually identify the size of the AS that hosts an IP infected by the worm by zooming in the map. Figure 9 shows a zoom operation near node 7018 (large AS).

The propagation analysis leads to the following observations: IMap shows a much clearer propagation pattern compared with a physical geographic map [5]. The worm spread to several ASes within the US within a very short time of its existence. This is an indicator that US hosts were the prime target of the worm. The worm quickly propagated to countries with strong (*i.e.*, many) connections to the US—Canada, Germany, France, Spain, UK, and Japan. It then infected neighboring countries of those in the second group

[10] A video demonstrating the spread per minute is available at the project website `http://vizsec-gama.cs.arizona.edu`.

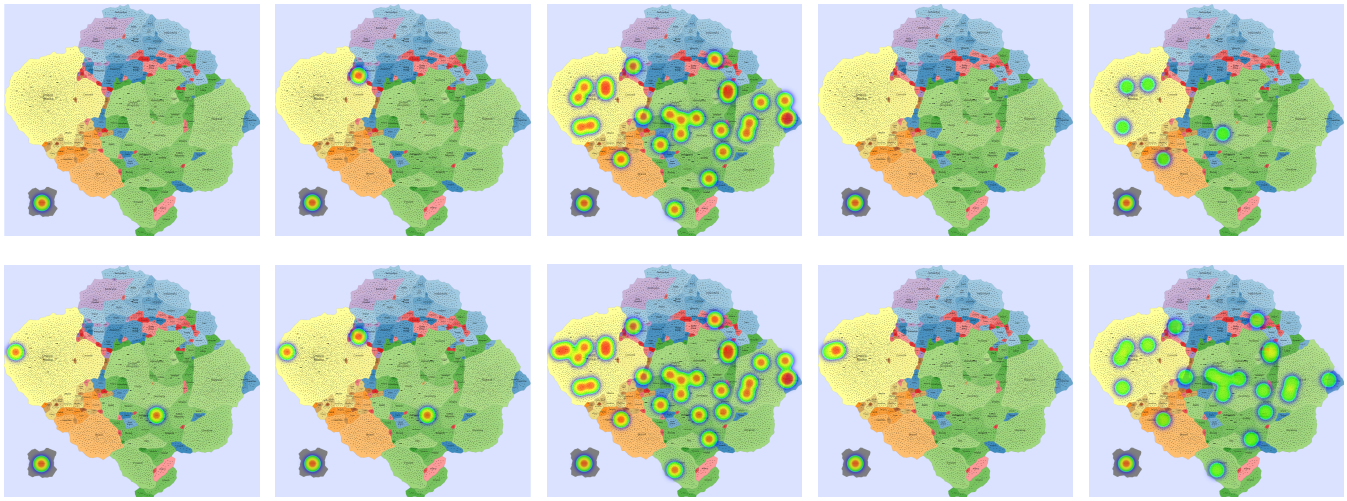
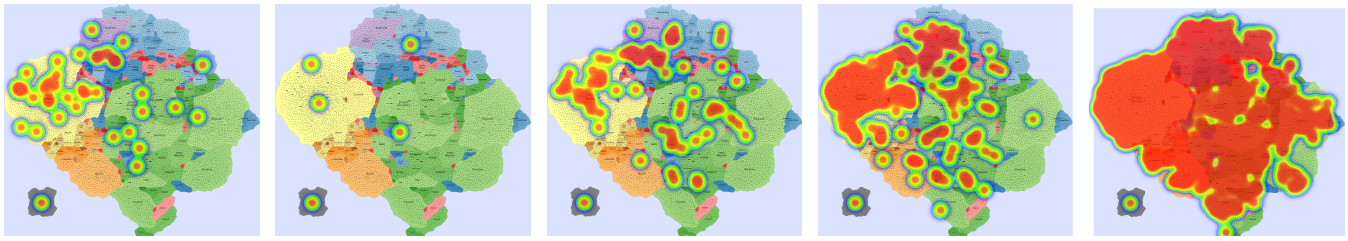


Figure 7: (DDoS) From left to right: Intervals 1 to 5 with heat maps generated with different metrics. First line shows results for metric C_1, while the second line shows results for metric C_2.



(a) Minute 1 (b) Minute 5 (c) 10^{th} hour (d) 13^{rd} hour (e) 17^{th} hour

Figure 8: Heat maps for Code-Red worm propagation. (a,b) Worm initially spreads; (c) reaches several countries at once; (d) propagates to several ASes within the same country; and then (e) attains the peak of activity.

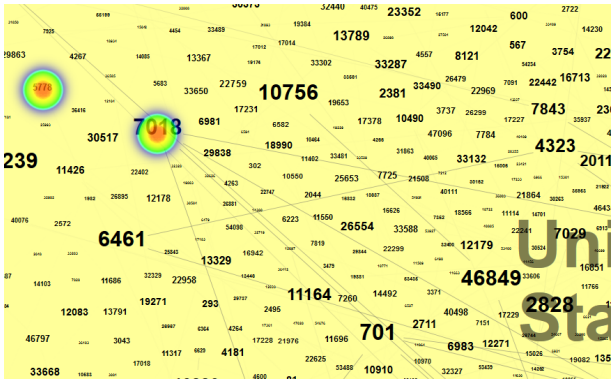


Figure 9: Heat maps showing node 7018 with edges between heavier nodes

(US → neighbors → neighbors).

Worm activity mostly started from small and medium nodes with several nodes from unknown sources. The *Unknown* country is highlighted in almost every interval (the number of unknown elements was checked in the pre-map processing step). In France, the worm spread from medium and small nodes in the first hours, until it reached an AS with a large weight in the 11^{th} hour, then the worm spread to many ASes within the country. In Germany, the activity was restricted to a group of small and medium-weight ASes until the 14^{th} hour, when it spread to other several small and medium ASes. The spread started at a rate $\lambda \leq 2.27$ nodes/minute until the 10^{th} hour, when the speed increased to $\lambda \leq 65.77$ nodes/minute. At that time the worm had affected most countries except for a few Asian countries, most countries in Africa and Central America. The unaffected countries host relatively few ASes, where a small number of ASes have sparse connections. However, unaffected regions might not have hosted vulnerable services or platforms affected by the worm. In some countries, the spread "exploded" upon reaching an AS with higher weight; in other countries the worm propagated only through small and medium ASes.

6. CONCLUSIONS AND FUTURE WORK

We proposed a novel visualization technique, IMap, which can be used for monitoring the security posture of a network of interest. In IMap, the Internet topology at the AS level is represented by a canonical map which resembles the geographic map of the world. The area, boundaries, and relative positions of IMap countries represent AS attributes and AS relationships of the Internet topology. To visualize

live traffic streams, aggregated IP traffic can be superimposed in the form of continuously updated heat maps. Heat maps aid security analysts in visually identifying the origin and magnitude of a security threats. We showcased two case studies, a synthetic DDoS attack and a worm propagation attack, to demonstrate how the intuitive and familiar to humans map metaphor facilitates the visualization, detection, and analysis of serious network anomalies.

As future work, we will use the intuitive geographic map metaphor to study the evolution of the Internet topology. Using a series of canonical maps, we will investigate the Internet's past evolutionary patterns and attempt to predict future ones. Moreover, we will study the impact of hypothetical catastrophic scenarios (elimination of critical nodes, links, and whole countries) on the Internet topology, connectivity, and overall performance.

Acknowledgments

This research was supported in part by the Office of Naval Research under Contract N00014-11-D-0033/0002. We would like to thank the Naval Research Laboratory and Ephibian, Inc for their valuable comments.

7. REFERENCES

[1] S. Bateman, R. Mandryk, C. Gutwin, A. Genest, D. McDine, and C. Brooks. Useful junk? the effects of visual embellishment on comprehension and memorability of charts. In *CHI*, pages 2573–2582, 2010.

[2] R. Blue, C. Dunne, A. Fuchs, K. King, and A. Schulman. Visualizing real-time network resource usage. In *VIZSEC*, pages 119–135, 2008.

[3] K. Boitmanis, U. Brandes, and C. Pich. Visualizing internet evolution on the autonomous systems level. In *Proc. of Graph Drawing*, 2008.

[4] R. Borgo, A. Adul-Rahman, F. Mohamed, W. p. Grant, I. Reppa, L. Floridi, and M. Chen. An empirical study on using visual embellishments in visualization. In *IEEE TVCG (InfoVis '12)*, pages 2759–2768, 2012.

[5] J. Brown. Animations for code-red worms spread, 2001. `http://www.caida.org/research/security/code-red/coderedv2_analysis.xml#animations`.

[6] CAIDA. Dataset on the code-red worms, 2001. `http://www.caida.org/data/passive/codered_worms_dataset.xml`.

[7] CAIDA. Visualizing internet topology at a macroscopic scale, 2005. `http://caida.org/analysis/topology/ascorenetwork`.

[8] CAIDA. UCSD IPv4 routed /24 topology dataset, month of july, 2014. `http://www.caida.org/data/active/ipv4_routed_24_topology_dataset.xml`.

[9] C. Collins, G. Penn, and S. Carpendale. Bubble sets: Revealing set relations with isocontours over existing visualizations. *IEEE TVCG*, 15(6):1009–1016, 2009.

[10] G. Di Battista, P. Eades, R. Tamassia, and I. G. Tollis. *Graph Drawing; Algorithms for the Visualization of Graphs*. Prentice Hall, July 1998.

[11] E. R. Gansner, Y. Hu, and S. G. Kobourov. Visualizing Graphs and Clusters as Maps. In *IEEE CGA*, pages 2259–2267, 2010.

[12] S. Ghani and N. Elmqvist. Improving revisitation in graphs through static spatial features. In *Proc. of Graphics Interface*, pages 175–182, 2011.

[13] J. R. Goodall, W. G. Lutters, P. Rheingans, and A. Komlodi. Focusing on context in network traffic analysis. *IEEE Comput. Graph. Appl.*, 26(2):72–80, Mar. 2006.

[14] R. Jianu, A. Rusu, Y. Hu, and D. Taggart. How to Display Group Information on Node–Link Diagrams: an Evaluation. *IEEE TVCG, To Appear*, 2014.

[15] T. Johnson and L. Lazos. Network anomaly detection using autonomous system flow aggregates. In *Proc. of GLOBECOM, to appear*, Dec 2014.

[16] C. Kintzel, J. Fuchs, and F. Mansmann. Monitoring large ip spaces with clockview. In *VIZSEC*, 2011.

[17] S. G. Kobourov, S. Pupyrev, and P. Simonetto. Visualizing graphs as maps with contiguous regions. In *EuroVis Short Papers*, pages 31–35, 2014.

[18] A. Lakhina, M. Crovella, and C. Diot. Characterization of network-wide anomalies in traffic flows. In *Proc. of SIGCOMM*, pages 201–206, 2004.

[19] M. Luckie, B. Huffaker, k. Claffy, A. Dhamdhere, and V. Giotsas. AS relationships, customer cones, and validation. In *IMC'13*, pages 243–256, 2013.

[20] P. Maigron. Regional internet registries statistics, 2014. `http://www-public.it-sudparis.eu/~maigron/RIR_Stats/RIR_Delegations/World/ASN-ByNb.html`.

[21] J. Oberheide, M. Karir, and D. Blazakis. Vast: visualizing autonomous system topology. In *VIZSEC*, 2006.

[22] R. Oliveira, M. Lad, and L. Zhang. Visualizing internet topology dynamics with cyclops, 2005.

[23] Y. Rekhter, T. Li, and S. Hares. RFC 4271: Border gateway protocol 4, 2006.

[24] F. Roveta, G. Caviglia, L. Di Mario, S. Zanero, F. Maggi, and P. Ciuccarelli. BURN: baring unknown rogue networks. In *VIZSEC*, 2011.

[25] B. Saket, P. Simonetto, S. Kobourov, and K. Börner. Node, node-link, and node-link-group diagrams: An evaluation. *IEEE TVCG (InfoVis14), To Appear*, 2014.

[26] P. Simonetto, D. Auber, and D. Archambault. Fully automatic visualisation of overlapping sets. *Computer Graphics Forum (EuroVis09)*, 28(3):967–974, 2009.

[27] G. Stapleton, P. Rodgers, J. Howse, and J. Taylor. Properties of Euler diagrams. In *Layout of Software Engineering Diagrams*, pages 2–16, 2007.

[28] M. Tory, D. W. Sprague, F. Wu, W. Y. So, and T. Munzner. Spatialization design: Comparing points and landscapes. *IEEE TVCG*, 13(6):1262–1269, 2007.

[29] W. Z. Yang Xiang, Ke Li. Low-rate DDoS attacks detection and traceback by using new information metrics. *IEEE Trans. on Information Forensics and Security*, 6(2), 2011.

[30] F. Zhou, R. Shi, Y. Zhao, Y. Huang, and X. Liang. Netsecradar: A visualization system for network security situational awareness. In *Cyberspace Safety and Security*, pages 403–416. Springer, 2013.