VisTracer: A Visual Analytics Tool to Investigate Routing Anomalies in Traceroutes

Fabian Fischer
University of Konstanz
Fabian.Fischer@uni-konstanz.de

Johannes Fuchs
University of Konstanz
Fuchs@dbvis.inf.uni-konstanz.de

Pierre-Antoine Vervier
Institut Eurecom
Pierre-Antoine.Vervier@eurecom.fr

Florian Mansmann
University of Konstanz
Florian.Mansmann@uni-konstanz.de

Olivier Thonnard
Symantec Research Labs
Olivier_Thonnard@symantec.com

ABSTRACT

Routing in the Internet is vulnerable to attacks due to the insecure design of the border gateway protocol (BGP). One possible exploitation of this insecure design is the hijacking of IP blocks. Such hijacked IP blocks can then be used to conduct malicious activities from seemingly legitimate IP addresses. In this study we actively trace and monitor the routes to spam sources over several consecutive days after having received a spam message from such a source. However, the real challenge is to distinguish between legitimate routing changes and those ones that are related to systematic misuse in so-called spam campaigns. To combine the strengths of human judgement and computational efficiency, we thus present a novel visual analytics tool named VisTracer in this paper. This tool represents analysis results of our anomaly detection algorithms on large traceroute data sets with the help of several scalable representations to support the analyst to explore, identify and analyze suspicious events and their relations to malicious activities. In particular, pixel-based visualization techniques, novel glyph-based summary representations and a combination of temporal glyphs in a graph representation are used to give an overview of route changes to specific destinations over time. To evaluate our tool, real-world case studies demonstrate the usage of VisTracer in practice on large-scale data sets.

Categories and Subject Descriptors

C.2.0 [**Computer-Communication Networks**]: General—*Security and protection*; C.3.8 [**Computer Graphics**]: Application; H.5.2 [**Information Interfaces and Presentation**]: User Interfaces

General Terms

Network Security, Visual Analytics, Traceroutes, Anomalies

VizSec '12, October 15 2012, Seattle, WA, USA

1. INTRODUCTION

Routing is a fundamental concept in the Internet. Correct path announcements are important to reach the correct destination servers. Despite of the importance and the severe consequences of routing issues, the responsible border gateway protocol (BGP) is quite vulnerable. Announcing malicious routing paths can be used to hijack IP blocks. As a result the attacker can conduct malicious activities from legitimate IP addresses. Distribution of vast amounts of spam is a scenario where the misuse of legitimate IP prefixes helps the attackers to circumvent widely used IP-based blacklists. The focus of this work is the large-scale analysis and exploration of routing anomalies for IP addresses starting to send spam in the Internet. This is achieved by actively tracking and measuring the traceroutes to the origin IP addresses over longer periods of time to eventually monitor possibly malicious path changes. Because of the vast amount of trace data with their changing underlying BGP routes, it is not helpful to just visualize the raw data. To make sense of the data it is important to algorithmically identify anomalies first. The tight integration of visual displays can be used to get an overview for quick ad-hoc analysis to identify noteworthy events and to differentiate them from false positives. The proposed visualizations in our work help to gain deep insights and visually explore the events within their context of historic and related anomalous traceroutes. Furthermore the analysts can push their findings back to the system. This feedback could then be used for further improving the underlying anomaly detection algorithms.

The three main contributions of our work are (i) a visual analytics tool called VisTracer to analyze large-scale traceroute data, (ii) the integration into our large-scale automatic analysis system and (iii) novel glyph- and graph-based summary visualizations for traceroutes. Additionally, we present an in-depth discussion of recent case studies for suspicious routing anomalies with respect to spam activities. The remainder of the paper is organized as follows. In Section 2 we discuss the most relevant related work. In Section 3 we describe the data infrastructure and the anomaly detection of our analysis approach. In Section 4 we present the proposed visual analytics tool and discuss real case studies in Section 5 to evaluate the system. Section 6 concludes with a summary and future work.

2. RELATED WORK

There are many tools, which visualize raw BGP update messages. Because these do also reflect the routing changes, such tools are highly related to our approach to visualize traceroutes. A popular approach to visualize such update messages are animated graphs, which can be seen in *BGPlay* [4], *LinkRank* [8] and *TAMP* [20]. Additionally, temporal information is presented as line charts, to explore and guide the animation of the path changes. Different paths are colored to enhance the readability or encode additional information. This helps to visually understand route withdrawals or update announcements. *BGPEye* [17] and [18] provide additional statistical information in a different view to monitor real-time routing activity like the moving average of the total number of BGP events or the deviation from historical trends. A combination of visualizing the Internet topology as a graph in a 3D display can be found in *VAST* [9]. *Elisha* [16] also uses a 3D display in combination with a pixel-based approach. Single pixels represent IP addresses and are colored according to previously classified BGP events. This space filling approach is enhanced by zooming and filtering techniques for gaining additional details. Again animation is used to visualize changes over time to be able to detect routing anomalies and MOASes. *BGPeep* [12] uses a more IP-space centered view in contrast to the previous approaches. Interesting ASes are visualized with a visualization similar to a parallel coordinates plot. One axis represents the AS and the other four the different octets of the IP address. A line traversing through the axes shows single update messages. There are also several commercial tools available related to visualizing BGP paths. RIPE provides visualizations for investigating BGP update messages. Similar to *BGPlay* their web-based tool *BGPviz*[1] visualizes routing changes as animated graphs in combination with a timeline. *RIPEstat*[2] offers various widgets with charts, tables and geographic maps to communicate detailed information about specific ASes. Statistical information enriches the data to better compare the current status with the historic activity. There is also a variety of tools visually representing single traceroutes as lines, either on a time axis or mapped on top to a geographic map. These visual tracerouting tools directly represent executed traces to single destinations.

Our approach instead makes use of a combination of pixel-based techniques to present anomalous events in an overview and glyph-based techniques to represent historical information for analyzed targets. We do also include a graph representation. However, our focus is the direct integration of temporal information into the nodes of the graph using a temporal glyph representation. Besides of the optional animation, this static integration in our approach can help the analyst to get a quick overview of the path without having to replay the whole communication as animation to understand the temporal changes.

Besides the visualizations of routing data, several solutions to secure BGP have already been studied in [3, 6] but the high computational cost of using cryptography and the required changes in the protocol and the infrastructure retain their deployment. BGP hijack detection techniques attempt to uncover abnormal changes in the routing infrastructure likely due to a BGP hijack by monitoring the *control plane* and/or the *data plane*. Methods described in [7, 10] only monitor BGP updates and trigger an alert when a new advertisement conflicts with their model of the Internet topology. In [2, 5, 21, 22, 15] data plane information is also used to collect information about the different hosts and networks along the forwarding path from a source to a monitored network. Several features of data plane traces can be leveraged to help detect abnormal routing changes, e.g., a network reachability change [15], an AS-level traceroute deviation [21], a significant change in the traceroute path length [22]. Finally, in [5] they combine control plane BGP hijack detection techniques with host fingerprints.

In our system we leverage different features of the traceroutes like the IP/AS paths, the route length, the host and AS reachability as well as some BGP information to detect abnormal routing changes. We also correlate them to help determine whether observed routing changes are benign or malicious.

3. DATA INFRASTRUCTURE

Manipulating the Internet routing infrastructure to hijack a block of IP addresses involves modifying the route taken by data packets so that they reach the physical network of the attacker. A system called SPAMTRACER [19] has been developed to monitor the routes towards malicious hosts by performing `traceroute` measurements repeatedly for a certain period of time. IP-level routes are translated into AS-level routes using live BGP feeds. The motivation for monitoring data plane routes towards specific hosts involved in spam campaigns is to collect the route taken by data packets to reach these hosts as soon as a spam is received from them. By performing multiple measurements on consecutive days for a certain period of time, typically one week, routes towards a given host or network can be compared and analyzed in depth to find evidences of a possible manipulation by an attacker of the routing infrastructure.

This system is based on a linear data flow where a feed of IP addresses to monitor is given as input and a series of enriched traceroute paths produced as output from which abnormal patterns can be uncovered. The incoming feed of IP addresses are retrieved from Symantec.cloud[3] spamtraps. This data is enriched with IP traceroutes. A customized version of the classic `traceroute` function is implemented and takes advantage of ICMP, UDP and TCP packets to increase the likelihood of hosts to be reached by them. Due to the many artifacts that can be found in IP-level traces, we also build the AS-level routes. The IP-to-AS mapping is performed using live and distributed BGP feeds from RouteViews [4] to obtain as accurate and complete mappings as possible. Additionally, information about the different hosts, AS owners, IP networks and geo locations is collected.

3.1 Extracting Routing Anomalies

We analyze the collected routes to uncover abnormal routing changes and classify them as benign or malicious. Routing anomalies are extracted independently for every monitored IP addresses. The first approach does focus on extracting routing anomalies from BGP hijacking scenarios, while the second one searches for suspicious patterns based on different metrics.

[1] http://www.ris.ripe.net/bgpviz/
[2] https://stat.ripe.net/
[3] http://www.symanteccloud.com/
[4] http://www.routeviews.org/

To identify malicious BGP hijacks, we start from existing scenarios of BGP hijacking [5] for which we know the resulting routing anomalies. However, it has to be considered that such routing anomalies can also result from benign BGP routing practices, e.g., multi-homing of customer ASes by ISPs, or from non-malicious incidents due to misconfiguration or operational errors. **Prefix Ownership Conflicts** occur when a block of IP addresses appears in the Internet routing infrastructure as originated by multiple ASes. This routing behavior can be the result of a hijacker advertising someone else's IP space in order to attract traffic to or originate traffic from that IP space. Advertising the *same prefix* is a possible way for BGP hijacking, if the IP prefix is already advertised by a different AS. This technique creates a routing anomaly referred to as Multiple Origin AS (MOAS). Announcing a slightly *different prefix* can also be used for tampering the ownership of a given IP prefix, which can be more (resp. longer) or less specific (resp. shorter). In this case, we refer to this anomaly as a Sub Multiple Origin AS (subMOAS). **BGP AS Path Anomalies** occur, when the location of a network in the Internet AS topology changes. As a result of a BGP hijack it is likely that the sequence of ASes traversed from two different points will change. Significant changes in the BGP AS paths should be investigated to determine if they are indeed benign or if they result from a malicious manipulation of the routing infrastructure. The *Next-Hop AS* anomaly can be observed with a certain number of different next-hop ASes, i.e., ASes next to the origin AS in an AS path, for a given origin AS and BGP collector. A *Complete AS Path* anomaly consists in observing a significant change in the AS paths for a given origin AS and BGP collector.

The second approach searches for suspicious patterns in traceroutes based mostly on metrics already used in previous works [22, 15]. **Traceroute Destination Anomalies** refer to suspicious values in features related to traceroute metadata. *Host/AS reachability* defines if a destination host or AS towards a given IP address is reachable (unreachable) for a certain number of days during the monitoring period and suddenly becomes unreachable (reachable) and remains like this until the end of the monitoring period. This reachability anomaly can result from a major routing change which causes the destination host or AS to become (un)reachable. The *hop count* or the length of a traceroute path is the value of the last TTL for which a reply to our probe IP packets has been received. The hop count anomaly is the consequence of a significant and sudden change in the hop count. This situation suggests that an important routing change occurred to permanently change the route taken by packets to reach the destination network. **Traceroute Path Anomalies** refer to suspicious changes in the sequence of hops traversed by traceroute paths to a given destination host. Using the different features collected for IP/AS hops, we can consider a traceroute not only as a sequence of IP addresses or ASes, but also as a sequence of countries, domain names, RIRs, etc. These alternate paths are leveraged in this detection of suspicious traceroute paths. The *AS-level Path Anomaly* consists in observing a significant change in the AS-level paths towards a given IP address. *Country-level Path Anomalies* are observed by extracting traceroute paths towards a given host exhibiting significant discrepancies in the sequence of traversed countries. This assumes that the countries traversed to reach a given destination from a given source is likely to remain constant even if routing changes occur at the IP or AS levels.

4. VISTRACER

The continuously growing SPAMTRACER database can be accessed by the analyst using our visual exploration tool called VISTRACER. The graphical user interface is built in a way to satisfy the needs of experienced analysts by providing an overview linked to more detailed visualizations. This helps to solve the different analysis tasks. The individual views can be placed according to the user's preference or adjusted to the working environment which is important, when the tool is used in multi-display environments.

The general workflow of VISTRACER is inspired by Shneiderman's information seeking mantra of having the overview first and then focusing on certain areas of interest to retrieve additional details [13]. The overall graphical user interface is shown in Figure 1. The left panel (1) provides a tabular anomaly view with all occurred anomalies. To investigate specific cases a filter box is integrated for quick ad-hoc queries. Using different constraints (2) for anomaly types and subtypes the user can focus on the different classes and combinations of anomalies. Based on the given constraints the *ASN Overview* (3) provides an overview of all anomalies using a visual representation. Findings can be stored in the database using the feedback panel (4), which can be used to annotate anomalies and comment on findings to make them accessible for other analysts. The right panel (5) provides tabular access to all destination targets with their traceroutes. Selecting entries in any of the tables will update the loaded visualizations for further investigation. A zoomable geographic map (6) to visually present the currently selected AS path is included. The *Visual Traceroute Summary* (7) is a compact visual representation, while the target graph visualization (8) can be used to get an in-depth overview of the temporal connections based on a graph-based approach.

4.1 ASN Overview

The main starting point for an exploratory analysis is to monitor different ASes and the occurring anomalies over time. Therefore, a zoomable matrix layout has been chosen as the basis for the visual marks shown in Figure 1 (3). The x-axis encodes the time and the y-axis the different destination ASes of traceroutes. By default, the ASes are ordered according to the total number of anomalies, while other sorting algorithms might be more appropriate for finding common patterns and correlations between different ASes. Due to the fact that multiple anomalies of *different* types can occur on specific points in time, rectangular glyphs are used to encode this additional information. Glyphs have the advantage of showing multiple data dimensions in a space efficient compact way. Each glyph has a fixed size and consists of several colored vertical stripes. Each colored stripe encodes one type of anomaly. The stripe width is proportional to the amount of daily anomalies for the respective event type. We decided to chose this additional size encoding to emphasize on the most prominent anomaly types in the overview, especially when they spread over longer periods of time. The stripe's color encoding is based on a qualitative color scale provided by Colorbrewer[5] and helps to visually distinguish

[5]http://colorbrewer2.org/

Figure 1: Graphical user interface of the VISTRACER visual analytics tool. (1) and (2) provide access to constraint filters and a table with observed anomalies. (3) Visual ASN Overview with occurred anomalies. A Feedback Panel is provided in (4) and access to individual traceroutes in (5) with map-based (6), glyph-based (7) and graph-based (8) visualizations.

between the different kinds of anomalies. Therefore, ASes with characteristic colored patterns are a visual hint for reoccurring anomalies. To further focus on the "hot spots" with lots of anomalies, opacity is used to encode the overall number of occurred events. Figure 2 (a) shows a closeup of such a single anomaly glyph.

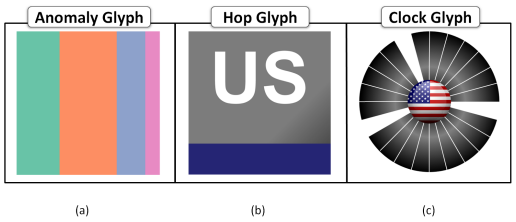

Figure 2: Three glyphs used in the visualizations.

An AS-based normalization is used to avoid artificially promoting heavily used large ASes. Suspicious ASes can be further investigated through double clicking on the different rectangles, which updates the different views and tables to provide more details on demand.

4.2 Target History Visualization

Traceroutes to the same destination can be investigated in the *Target History Visualization*. The main idea of this visualization is to provide a visual traceroute summary to show hop usage variances of single traceroutes to the same target. Therefore, the x-axis encodes the individual hops and the y-axis the traceroutes on the different days. Whenever a hop is used within a traceroute a small glyph is placed accordingly. This rectangular glyph encodes the country code of the hop with a small label and a colored bar. With the help of this colored bar, connections within the same country can be spotted preattentively. The main color of the glyph reflects whether the traceroute was complete (green) or incomplete (gray). This prominent feature is visible at first sight because it is considered of high importance. Additionally, brightness is used to encode the latency of the individual hops. A closeup of this glyph can be seen in Figure 2 (b). At the end of each traceroute row, a small anomaly container is placed. The container represents the four main types of anomalies with equally sized rectangles. These rectangles are further divided into smaller rectangles representing the subtypes. Whenever a type/subtype combination can be found in a traceroute the corresponding rectangle is colored. Thus, anomalies lasting for a longer period can be easily detected as a reoccurring pattern over many traceroutes. Suspicious traceroutes with lots of anomalies show several colored rectangles and, therefore, are easy to spot. Examining the anomalies in combination with the used hops and the completeness of the traceroutes over time can lead to relevant findings and helps the analyst to understand the traceroutes. This visualization is especially effective to get an overview of the used hops in the different traceroutes.

4.3 Temporal Graph Representation

The previous visualization does not focus on following the exactly used routes or the identification of the most common route in the correct order. To solve this task, an additional graph visualization is provided. The graph layout is extended with an additional glyph encoding to show routing changes over time. The nodes represent the different hops, the edges show the connections with each other. The width of an edge depends on the amount of traces using this exact connection. The nodes are visualized by circular glyphs with equally sized slices and small flags reflecting the country of the hop as can be seen in Figure 2 (c). Because of the aspect ratio, the circular glyphs can directly be integrated into the graph nodes without wasting additional space for this temporal information or requiring disturbing and more time-consuming animation. The number of slices depends on the amount of traceroutes shown in the graph. The clockwise arranged slices represent the different traceroutes for the selected days. When a hop was used in a traceroute the respective slice is colored, otherwise it is not displayed at all. The color depends on whether the traceroute reaches its destination or not. This encoding supports the analyst in detecting the main route (i.e., based on the path's width), the usage of hops (i.e., the proportion of colored slices), the reachability of the destination (i.e., the hue of the colored slices) and the temporal development of the route (i.e., the partition of the slices). Additionally, the geographic location of the corresponding country can be taken into account in the layout to highlight possible route flappings between different countries with the help of the graph's layout. To focus on the main route, we additionally propose an *Enhanced Baseline Layout* which displays the most common path at the bottom. The hops, not being part of the baseline are arranged in a force-directed way above the baseline.

Combining the different views or looking at them individually supports the user in the different analysis tasks. To evaluate the tool's effectiveness, the following section describes the analyst's workflow and how the visualizations help.

5. CASE STUDIES

In this section we describe how suspicious routing events are identified and how the VISTRACER framework reflects this workflow to assist the analyst. We also present two case studies of routing events identified as suspicious using the developed visualization tool.

5.1 Visual Analysis Workflow

Figure 3 depicts the steps involved in the analysis of the network traces collected by SPAMTRACER. Furthermore, this figure shows where in the workflow the visualizations can assist the analyst in examining the data. In detail the analysis is based on (i) automatically extracting routing anomalies from the traces as described in Section 3, (ii) selecting the monitored hosts having a meaningful set of anomalies, and (iii) investigating cases using all the collected data to identify the suspicious cases. The result of the investigation of a case is finally reported back to the database (iv).

VISTRACER supports the *Selection of Candidate Suspicious Cases* by providing a graphical user interface to filter for anomalies which match a given set of constraints on the type, the number and the time of appearance of the anomalies. These correspond to the most likely suspicious cases.

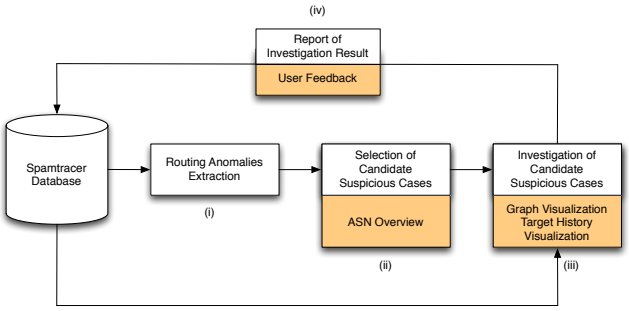

Figure 3: Overview of Visual Analysis Workflow.

This step is associated with the *ASN Overview Visualization*, which allows the analyst to define the constraints on the anomalies and then explore the resulting set of targets aggregated at the AS level. The *Investigation of Candidate Suspicious Cases* means to investigate the suspicious cases with the help of the collected traces as well as some external routing information services to determine if a case is benign or if it results from a malicious BGP hijack. When investigating a case, the *Graph* and *Target History Visualization* as well as the traceroute hop list provide the analyst with all the data available to determine whether the routing anomalies observed reflect a malicious routing behavior. To communicate and further make use of the findings the tool also focuses on *Reporting of Investigation Results*. The feedback loop embedded in VISTRACER allows to share the result of the investigation with other analysts.

The SPAMTRACER data set used to produce the two case studies contains traceroutes collected from April 2011 until the end of August 2011. 848,916 data plane routes were collected towards 239,907 IP addresses and 5,912 ASes. After the routing anomalies were extracted from the traces 41,430 destination IP addresses were found to have at least one anomaly. Given the high number of cases exhibiting at least one anomaly, we decided to focus on cases having the following combinations of anomalies:

- *BGP Origin & BGP or Traceroute Path Anomalies:* Select cases exhibiting a Prefix Ownership Conflict with a significant change in the BGP or Traceroute AS path.
- *BGP Origin & Traceroute Destination Anomalies:* Select cases exhibiting a Prefix Ownership Conflict with either an IP/AS reachability change or a significant data plane route length change.
- *Traceroute Destination Anomalies & BGP or Traceroute Path Anomalies:* Select cases exhibiting a significant change in the BGP or Traceroute AS Path with an IP/AS reachability change or a significant data plane route length change.

We have thus applied these filters in the *Traceroute Anomalies* panel of VISTRACER to focus our analysis on these cases.

5.2 Analysis of Suspicious BGP Anomaly

The first case study presents the visual analysis of a network whose traffic was apparently hijacked by another AS. Actually, we show how such a case can be uncovered and

investigated using the visualizations and other information provided by VISTRACER.

From the *ASN Overview* visualization, one particular case caught our attention, which can be seen in Figure 4. Two ASes actually appeared to share several anomalies, which occurred on the same day. The visualization allows to extract such time correlation between anomalies in different ASes thanks to the ASNs and time dimensions. Looking at the anomalies extracted for the two ASes reveals (i) a Traceroute Destination Anomaly (related to the destination AS reachability), (ii) Traceroute Path Anomalies, (iii) BGP Path Anomalies (AS Path Deviation) and, (iv) a BGP Origin Anomaly (related to a subMOAS conflict).

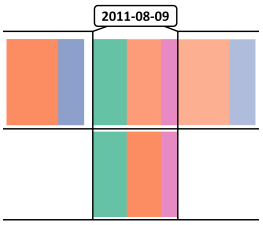

Figure 4: Closeup of the *ASN Overview* showing two nearly identical anomaly distributions for two different ASN at the same point in time.

We can make use of the *Target History Visualization* to have a first view of the traceroute paths and the uncovered routing anomalies. Figure 5 shows the *set* of IP hops traversed by traceroutes from the vantage point in France to the destination host throughout the monitoring period. From this visualization we can say that there is a noticeable change in the set of traversed IP hops between the third and the fourth traceroute. The six routing anomalies uncovered for these traceroutes on the fourth day confirm that a major routing change occurred. In this case, a change in the origin AS of the destination IP prefix occurred at the same time as a change in the sequence of ASes traversed both in the traceroutes and in the BGP AS paths. The BGP Origin Anomaly, in the third column, has been marked as benign (green) by SPAMTRACER, because the two conflicting ASes were found to have a provider-customer relationship.

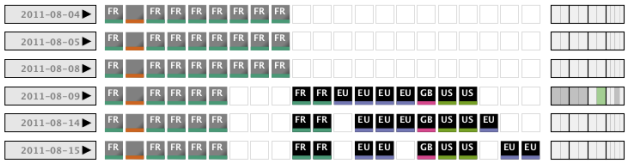

Figure 5: *Target History Visualization* of the first case study: The visualization shows the significant difference in the ASes traversed between the third and fourth day. The routing anomalies observed on the fourth day are also shown.

To further investigate the case, we make use of the *Graph Visualization*, which is presented in Figure 6 for the same monitored host. The *Graph Visualization* allows the analyst to look at the IP-, AS- or the Country-level traceroute paths, i.e., the *sequence* of IP hops, ASes or countries traversed. While the AS-level graph is particularly well suited to investigate abnormal changes in inter-domain routing, the IP- or Country-level graphs can also be leveraged to investigate routing anomalies. Actually, they are complementary. It is thus interesting to start from the high-level view of the Country-level graph and go down the levels to analyze in more details specific parts of the routes.

In the present case we decide to make use of the AS-level graph to compare the sequence of traversed ASes before and after the change of origin AS. The origin and destination AS before the change belongs to a backbone ISP, which advertises an aggregated IP prefix including the destination IP prefix. The unreachability of the destination AS after the change can be observed on day four and correlated with the Traceroute Destination Anomaly seen on the same day in the *Target History Visualization*. Also, the last AS that could be reached by traceroutes appears in the collected BGP AS paths, as the next-hop AS, which is the direct upstream provider, of the new origin AS. This provider-customer relationship could not be officially explained. Hijacking a network can actually be performed by advertising it with a correct origin AS and by putting the attacking AS as the next-hop AS.

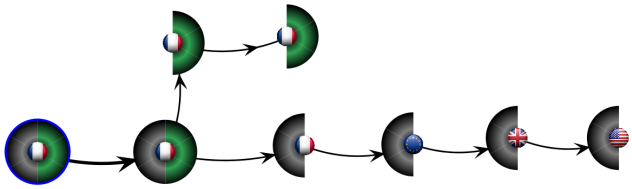

Figure 6: The *Graph Visualization* shows the significant difference in the ASes traversed between the third and fourth day.

After an investigation, it turned out that the next-hop AS belonged to a company providing DDoS mitigation as service by sink holing the attacking traffic of their customers. The analysis suggests that either the security company redirected the traffic of their customer's AS because they were under attack or the security company may sometimes act as an ISP for some companies' AS to easily protect them from undesired traffic. Given the fact that the security company advertised the route in BGP for at least three days, we believe that it actually acted as an ISP for its customer.

Although we have detected abnormal routing changes regarding this network, it is quite difficult to validate these anomalies as a real hijack case since we lack the feedback from the owner of the network.

5.3 Link Telecom BGP Hijack

This second case study presents the visual analysis of a *validated* BGP hijack performed by a spammer to send spam from the stolen IP address space. The hijacking spammer phenomenon has already been observed in [11, 5] and consists of spammers taking control of unused IP address space in order to send spam from clean, non-blacklisted IP addresses.

From the *ASN Overview* (Figure 7), AS31733 caught our attention, because many diverse routing anomalies occurred within a limited period of time. Moreover, several anomalies occurred on the same day, which reinforced the idea that a major routing change occurred at that time for this AS. The

uncovered anomalies related to AS31733 include (i) Traceroute Destination Anomalies (related to the destination host and AS reachability), (ii) Traceroute Path Anomalies and, (iii) BGP AS Path Anomalies (AS Path Deviation).

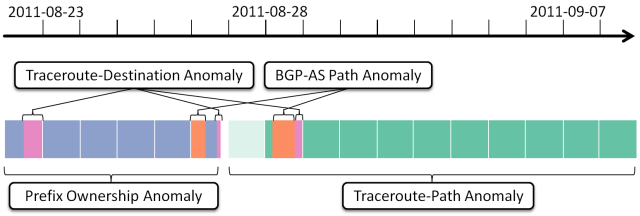

Figure 7: The *ASN Overview* of AS31733 reveals many different anomalies over a longer period of time.

The *Target History Visualization* of a monitored host within AS31733 exhibiting a combination of Traceroute Destination Anomalies, Traceroute Path Anomalies and BGP AS Path Anomalies. Figure 8 presents the *Target History Visualization* which shows the *set* of ASes traversed by traceroutes from the vantage point in France to AS31733 throughout the monitoring period. We can clearly see that the set of traversed ASes changes significantly. By looking at the anomalies extracted for that case, we can also see that all anomalies were observed on a particular day, i.e., just after the change in the traceroute path. The observation of the set of IP hosts traversed by the traceroutes shows the exact same behavior. From these observations we can say that the location of the monitored AS in the Internet AS topology changed significantly.

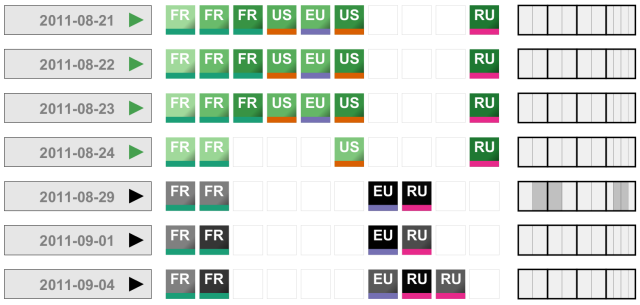

Figure 8: The *Target History Visualization* shows the significant difference in the *set* of ASes traversed between the fourth and fifth day. The routing anomalies observed are also shown.

Figure 9 presents the *Graph Visualization* of the same monitored host within AS31733. This visualization shows the *sequence* of IP hops, ASes or countries traversed by the traceroutes. In this case, looking at the Country-level paths would show that packets always seem to go through the US to go from a source in France to a destination in Russia. While this routing behavior can be considered abnormal, we also know that some big ISPs, i.e., backbone ISPs, are spread across continents and may be introduce US hops in a European route. If we now look at the AS-level graph we can see that US ISPs Level-3 (AS3356) and Internap (AS12182) both appear in the routes. Besides being a backbone ISP, Level-3 also appears in every traceroute during the monitoring period. However, Internap only appears in the first traceroute, before the routing change. To have more details about the traceroute going through AS12182 Internap, we can have a look at the IP-level graph. The graph reveals that the first traceroute goes through two routers of AS12182 apparently located in the US and then directly ends in AS31733 apparently located in Russia. This suggests that the destination host currently using an IP of AS31733 is likely located in the US instead of Russia. Furthermore, the visualization also shows that the destination host and AS could not be reached from the fifth day until the end of the monitoring period. This observation is corroborated by the Traceroute Destination Anomalies (related to the host/AS reachability) uncovered on the fifth day. All this suggests that the routing change observed lead to the destination host and AS to become unreachable.

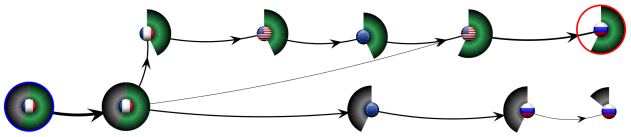

Figure 9: The *Graph Visualization* shows the significant difference in the *sequence* of ASes traversed. It also highlights the unreachability of the destination AS after the routing change occurred.

After the investigation, it turns out that on August 20^{th} 2011 the network administrator of the Russian telecommunication company "Link Telecom", whose AS31733 belongs to, complained on the North American Network Operators' Group (NANOG) mailing list that his network had been hijacked by a spammer [1]. On both August 25^{th} and August 29^{th} 2011 changes were observed in the traceroutes and BGP routes towards AS31733. These changes were the result of the owner regaining control over his network. In this case, the aggregation in the *ASN Overview* of the routing anomalies extracted for the individual monitored hosts within their AS actually uncovered the pattern of several diverse and timely close routing anomalies.

This hijack case is further described in [14]. Although the prefix appeared to be announced by the correct origin AS, i.e., AS31733, it was routed via a US ISP called Internap (AS12182). During this period the network was under the control of the spammer, spam messages were received by Symantec.cloud honeypots. The hijack lasted for five months from April 2011 until August 2011 and is a validated case of a hijacking spammer that managed to steal someone else's IP space and sent spam from it.

6. CONCLUSION

In this work we presented a novel visual analytics tool called VISTRACER to investigate routing anomalies and BGP hijacks. In particular, spamming activities were monitored with the help of a large-scale traceroute collection system. Special care was taken to design VISTRACER to support the workflow of analyzing the large-scale dataset according to the analysts' needs. The tool's flexibility is derived from the integration of several linked data views and visualizations into a powerful analysis suit, which can address a variety of analysis questions. Furthermore, the usefulness and ef-

fectiveness of VISTRACER for network security analysts was demonstrated in two case studies.

In the future we will integrate different additions to further improve the usability of the tool. Regular usage of VISTRACER by our analysts will also show, which additional views should be integrated. To improve the scalability of the graph representation, further layout improvements will be made to reduce possible clutter of traceroutes with very complex connections and to incorporate missing hops in the layout.

7. ACKNOWLEDGMENTS

The research leading to these results has received funding from the European Commission's Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 257495, "Visual Analytic Representation of Large Datasets for Enhancing Network Security" (VIS-SENSE).

8. REFERENCES

[1] Prefix hijacking by michael lindsay via internap. `http://mailman.nanog.org/pipermail/nanog/2011-August/039381.html`, August 2011.

[2] H. Ballani, P. Francis, and X. Zhang. A study of prefix hijacking and interception in the Internet. In *Proceedings of the 2007 conference on Applications, technologies, architectures, and protocols for computer communications*, SIGCOMM '07, pages 265–276, New York, NY, USA, 2007. ACM.

[3] R. Bush and R. Austein. The RPKI and Origin Validation, June 2009.

[4] L. Colitti, G. Di Battista, F. Mariani, M. Patrignani, and M. Pizzonia. Visualizing Interdomain Routing with BGPlay. *Journal of Graph Algorithms and Applications*, 9(1):117–148, 2005.

[5] X. Hu and Z. M. Mao. Accurate Real-Time Identification of IP Prefix Hijacking. In *Proceedings of the 2007 IEEE Symposium on Security and Privacy*, SP '07, pages 3–17, Washington, DC, USA, 2007. IEEE Computer Society.

[6] S. Kent. Securing the Border Gateway Protocol: A Status Update. In *In Seventh IFIP TC-6 TC-11 Conference on Communications and Multimedia Security*, pages 2–3, 2003.

[7] M. Lad, D. Massey, D. Pei, Y. Wu, B. Zhang, and L. Zhang. Phas: A prefix hijack alert system. In *Proc. USENIX Security Symposium*, 2006.

[8] M. Lad, D. Massey, and L. Zhang. Visualizing internet routing changes. *IEEE Transactions on Visualization and Computer Graphics*, pages 1450–1460, 2006.

[9] J. Oberheide, M. Karir, and D. Blazakis. VAST: visualizing autonomous system topology. In *Proceedings of the 3rd international workshop on Visualization for computer security*, pages 71–80. ACM, 2006.

[10] J. Qiu and L. Gao. Detecting bogus bgp route information: Going beyond prefix hijacking. Technical report, In Proc. SecureComm, 2007.

[11] A. Ramachandran and N. Feamster. Understanding the network-level behavior of spammers. In *SIGCOMM '06: Proceedings of the 2006 conference on Applications, technologies, architectures, and protocols for computer communications*, pages 291–302, New York, NY, USA, 2006. ACM.

[12] J. Shearer, K. Ma, and T. Kohlenberg. BGPeep: An IP-Space Centered View for Internet Routing Data. *Visualization for Computer Security*, pages 95–110, 2008.

[13] B. Shneiderman. The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations. In *Proceedings 1996 IEEE Symposium on Visual Languages*, pages 336–343. IEEE Computer Society, 1996.

[14] Symantec Corporation. Symantec Internet Security Threat Report. `http://www.symantec.com/threatreport/`, April 2012.

[15] M. Tahara, N. Tateishi, T. Oimatsu, and S. Majima. A Method to Detect Prefix Hijacking by Using Ping Tests. In *APNOMS '08: Proceedings of the 11th Asia-Pacific Symposium on Network Operations and Management*, pages 390–398, Berlin, Heidelberg, 2008. Springer-Verlag.

[16] S. T. Teoh, K. L. Ma, S. F. Wu, and X. Zhao. Case study: interactive visualization for internet security. In *Proceedings of the conference on Visualization '02*, VIS '02, pages 505–508, Washington, DC, USA, 2002. IEEE Computer Society.

[17] S. T. Teoh, S. Ranjan, A. Nucci, and C.-N. Chuah. Bgp eye: a new visualization tool for real-time detection and analysis of bgp anomalies. In *VizSEC '06: Proceedings of the 3rd international workshop on Visualization for computer security*, pages 81–90, New York, NY, USA, 2006. ACM.

[18] S. T. Teoh, K. Zhang, S.-M. Tseng, K.-L. Ma, and S. F. Wu. Combining visual and automated data mining for near-real-time anomaly detection and analysis in bgp. In *VizSEC/DMSEC '04: Proceedings of the 2004 ACM workshop on Visualization and data mining for computer security*, pages 35–44, New York, NY, USA, 2004. ACM.

[19] P.-A. Vervier and O. Thonnard. Spamtracer: Using Traceroute To Tracking Fly-By Spammers (under review). In *The 8th International Conference on emerging Networking EXperiments and Technologies*, CoNEXT '12, Nice, France, 2012. ACM.

[20] T. Wong and C. Alaettinoglu. Internet routing anomaly detection and visualization. In *Dependable Systems and Networks, 2005. DSN 2005. Proceedings. International Conference on*, pages 172–181. IEEE, 2005.

[21] Z. Zhang, Y. Zhang, Y. Charlie, H. Z. Morley, and M. R. Bush. iSPY: Detecting IP Prefix Hijacking on My Own. In *In Proc. ACM SIGCOMM*, 2008.

[22] C. Zheng, L. Ji, D. Pei, J. Wang, and P. Francis. A light-weight distributed scheme for detecting IP prefix hijacks in real-time. In *Proceedings of the 2007 conference on Applications, technologies, architectures, and protocols for computer communications*, SIGCOMM '07, pages 277–288, New York, NY, USA, 2007. ACM.