SEVENTH FRAMEWORK PROGRAMME
Area ICT-2009.1.4 (Trustworthy ICT)

VIS-SENSE

Visual Analytic Representation of Large Datasets for Enhancing Network Security

D6.2 BGP Analysis Scenario

Contract No. FP7-ICT-257495-VIS-SENSE

Workpackage	WP 6 - Workpackage 6
Author	EURECOM
Version	1
Date of delivery	M36
Actual Date of Delivery	M38
Dissemination level	Public
Responsible	EURECOM
Data included from	EUR, SYM, CERTH, UKON

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement n°257495.

SEVENTH FRAMEWORK PROGRAMME
Area ICT-2009.1.4 (Trustworthy ICT)

The VIS-SENSE Consortium consists of:

Fraunhofer IGD	Project coordinator	Germany
Institut Eurecom		France
Institut Telecom		France
Centre for Research and Technology Hellas		Greece
Symantec Ltd.		Ireland
Universität Konstanz		Germany

Contact information:
Dr Jörn Kohlhammer
Fraunhofer IGD
Fraunhoferstraße 5
64283 Darmstadt
Germany

e-mail: joern.kohlhammer@igd.fraunhofer.de
Phone: +49 6151 155 646

Contents

Abstract

This Deliverable completes Task 6.2 by evaluating the advancements made by the VIS-SENSE project to BGP hijacking detection. We take advantage of the VIS-SENSE framework developed during WP5 in order to isolate and analyze network events where the routing infrastructure was attacked in order to carry out other malicious activity on the network. We present the Link-Telecom hijack case, where an American spammer abused a Russian ISP's prefix during four months in 2011. We present cases where fly-by spammers abused IP blackspace in the first six months of 2013. We present a case where a spammer was believed to have hijacked a Bulgarian ISP's subprefixes in order to carry out spam and scam activities.

1 Introduction

The global goal of WP6 is to evaluate the VIS-SENSE framework. As described in the Description of Work, the goal of Task 6.2 is to focus on the evaluation of BGP advancements archived by the VIS-SENSE project through the use of the VIS-SENSE framework. By using advanced analysis techniques and visual analytics tools, routing events where attackers appear to have abused the routing infrastructure as a stepping stone to run a sophisticated attack can be highlighted. This Deliverable aims at reporting those events; and special care is taken in order to validate them (i.e. ensure that they are indeed the result of an attack, and not a false positive).

This Deliverable is structured as follows. Chapter 2 discusses the concepts and methods developed during VIS-SENSE WP3 and WP4. It describes the BGP analysis algorithms designed and developed by the VIS-SENSE consortium, and their related visualization techniques. Chapter 3 discusses the challenges faced by the analysis of BGP and BGP hijacking attacks. It underlines the weaknesses inherent to the available BGP data due to protocol design, and presents the reason why valid ground-truth information is inexistent or hard to obtain. In order to remedy this lack of ground-truth, several analysis methods for gaining insights into a suspicious case are presented. Chapter 4 uses the tools developed by the VIS-SENSE consortium in order to isolate and analyse a set of suspicious BGP events. The first part of the Chapter focuses on the Link-Telecom, a confirmed, long-term hijack in which an American spammer abused a Russian ISP's defunct prefix in order to emit spam. The second part of the Chapter describes a form of fly-by spammers uncovered by the VIS-SENSE framework. The last part of the Chapter analyses the Bulgarian case, where a spammer is suspected of having hijacked a set of a Bulgarian ISP's subprefixes, and then carried out spam and scam activities with these prefixes. Finally, Chapter 5 summarizes and concludes the advancements made by the VIS-SENSE consortium in the land of prefix hijacking detection. The rest of this Chapter serves as a reminder of the basic notions used in this Deliverable.

1.1 BGP and Prefix Hijacking

The Internet is composed of hundreds of thousands of independent Autonomous Systems (AS). Routing in the inter-AS domain is achieved by the unique protocol BGP (Border

Gateway Protocol), which was first defined in June 1989. The current version of the protocol, version 4, is defined in RFC4271 [35].

During normal BGP operations, two *peering* routers propagate reachability information through the exchange of *update messages*. These messages are composed of a list of (IP) prefixes to *withdraw*, and of another list of prefixes to *announce* along with their attributes. One such attribute is the *AS path*, to which each router propagating a route *prepends* its own, globally unique AS number. As a result, the *origin* of a route is always the rightmost AS number in the AS path (unless the route has been aggregated, in which case the AS path becomes unordered and is referred to as an *AS set*).

Just like many legacy protocols, BGP relies on mutual trust: any BGP-enabled router can announce any prefix to its peering BGP routers, leaving BGP unable to authenticate origins, paths, or topology. This lack of information renders prefix hijacking situations possible. A *prefix hijacking* is a situation where an AS originates routes to prefixes it does not own.

Hijacks can be the result of local router misconfigurations, such as on April 25, 1997 when a regional ISP in the USA leaked routes to a backbone provider [13, 18, 31]. As a result, the global network was blackholed for around 6 hours. Weaknesses of BGP gained global media attention on February 24, 2008, when Pakistan Telecom decided to block access to YouTube by BGP means [36]. Due to a misconfiguration, they started announcing a set of prefixes covering YouTube's IP range globally. For about 1h30mn, YouTube was effectively offline. On April 8, 2010, China Telecom leaked thousands of prefixes affecting networks owned by large international corporations, as well as foreign intelligence services [47, 26, 25, 29, 49]. Mainstream media qualified the event as cyber-war.

However, hijacks can also be the result of an attack on the routing infrastructure [34, 11, 22, 45, 14, 33, 42], through which an attacker can effectively steal the victim's network identity, or spy on its traffic. Ramachandran et al. [34] postulated the existence of so-called *fly-by spammers* that use spectral agility – a short-lived hijack of a large prefix – in order to stealthily emit spam in the network. This behaviour was also reported by Hu at al. [22]. More recently, we reported and analysed the Link-Telecom hijack case, where a group of spammer used a long-term hijack of black IP space[1] [12, 41].

Taxonomies of prefix hijacking attacks were presented in [27, 22]. *Ownership attacks* see an attacker originating a prefix as if they were the real owner. In an *AS path attack*, an attacker tampers with the AS path of an announcement, effectively creating so-called fake-edges in the network. *Subprefix attacks* take advantage of BGP always selecting the most specific prefix in order to route traffic. *Man-in-the-middle attacks* are

[1] i.e. assigned, but unannounced IP prefixes

possible with BGP, and take advantage of these three elements: the attacker originates subprefixes of the victim's and tampers with the AS path on these announcements so that the ASes located between the attacker and the victim remain unaffected by this announcement. This effectively reroutes the traffic destined to the victim from the Internet to the attacker, who can then forward the traffic back to the legitimate network through the unaffected original path[2] [32].

Prefix hijacking cannot be avoided because it is the result of believing in mutual trust at design time. Secure BGP (S-BGP) and Secure Origin BGP (soBGP) have been proposed as addenda to BGPv4. Both leverage public key infrastructure (PKI) certificates in order to authenticate the players in an issued route: origins, paths, and owners. Concerns have been raised over the feasibility of large-scale use of these protocols, because of their configuration complexity and increased convergence time [14]. More recently, Resource Public Key Infrastructure (RPKI) initiative has been gaining momentum. It is a community-driven dictionary of certificates that enable origin validation (i.e. that ensures an AS number is indeed authorized to originate a prefix) known as Route Origin Authorizations (ROAs) [46, 38]. However, so far, only around 4% of global routes are RPKI enabled [16].

Because there is no short-term possibility on preventing hijacking attacks, there has been a lot of focus on implementing techniques that detect occurrences of these attacks. Detection techniques are usually divided into two distinct categories: those based on the *control plane*, i.e. the detection signature depends on information found in a router's routing table and/or messages exchanged with this router; and those based on the *data plane*, i.e. the detection signature is based on the way packets actually flow between an observer and the source. RIPE RIS [5], among others, offers binary dumps of BGP messages exchanged by their routers, as well as snapshots of their routing tables to perform control plane analysis.

Detection techniques from the control plane, such as [27, 11, 33], involve the creation of a model that represents the normal, expected behaviour of a network. Whenever the current view of the network differs from the model, an alert is raised. The complexity and accuracy of the model is then the key element to a good detection scheme.

Detection techniques from the data plane, such as [52, 22, 45, 50, 21] involve active probing of the network topology and/or available live hosts in the monitored network. The core idea is that when a hijacking takes place, significant topology changes should be observed, while the victim network is different from the hijacking network. The way these elements are measured, as well as their diversity, ensures a good detection.

[2] Of course, the traffic between the victim and the networks located on the path between the attacker and the victim is not hijacked.

More recently, Argus [42] proposed an integrated system that leverages both approaches by running ping tests from a set of distributed vantage points when an anomaly is detected from the control plane. By using distributed vantage points, discrepancies, if any, between the polluted part of the Internet (i.e. the part of the Internet that received the suspect BGP route) and the regular part of the Internet (i.e. the part of the Internet that has *not* received the new, suspect BGP route). A correlation function is then used on the information received from both parts of the Internet in order to determine if the event is due to a hijack, a route migration, or standard routing practices.

Unfortunately, it is hard to assess the quality of these techniques because the signatures used to detect prefix hijacking also match standard routing/network engineering practices. For example, a MOAS (Multiple-Origin AS) – a situation in which a prefix is originated by multiple AS numbers – can be the result of an ownership attack, but is even more likely the result of a network engineering practice, such as anycasting, or multihoming [27]. AS path anomalies can be the result of an attack, but also, for example disaster recovery simulation, or downright underestimation of the AS-level links in the Internet [40]. As a result, the output of prefix hijacking detection algorithms are currently mostly populated with false positives, i.e. events that do not pose a security threat but are the result of standard routing practices.

1.2 VIS-SENSE Data Infrastructure

The VIS-SENSE project uses a diverse set of data sources, described in WP2's D2.2: Data Collection Infrastructure. This Section quickly summarizes the datasets that were used in order to underline the results presented in Chapter 4. As with every other VIS-SENSE dataset, a WAPI server was setup in order to share these dataset among the VIS-SENSE consortium.

1.2.1 BGPDB

The observation of the BGP control plane has to be done from inside a BGP-enabled router. Since the infrastructure necessary to access such a router is quite large, some entities, such as the Routing Information Service (RIS) [5] project from RIPE NCC [4] provides access to 13 geographically diverse *looking glasses*: routers that can be remotely accessed in order to view the current status of their routing tables. RIS also provides dump files of the messages exchanged with their BGP peers, as well as snapshots of their routing table. These dump files are created every five minutes and made available in compressed form at [30]. Consequently, in order to monitor BGP, one first has to download a large number of files and then parse them. This is particularly wasteful if one is only interested in studying the behaviour of a small number of prefixes (e.g. a manual lookup following an alert from a monitoring algorithm). Moreover, our experience is that the messages contained within these files are not necessarily chronologically ordered: it is not uncommon, while parsing the file, to stumble on a file whose timestamp is older than the one preceding it. This is problematic when trying to recreate the router's state at a given time: if the messages are not processed chronologically, the router's state cannot be accurately recreated.

BGPDB was designed to solve these problems. Its goal is to gather in a single, easily accessible location messages from RIPE RIS' collectors. Because it is a database, it is also easily searchable, meaning that only messages related to a given prefix can be downloaded; or that the ASNs that originated a given prefix can be returned immediately. Because it is sortable, a chronological order of messages can be guaranteed.

BGPDB has also been extended to contain processed results by MOAS filtering (the technique is detailed in Section 2.1.6), and thus provides security-related data (suspicious prefixes) along with analysis data (BGP update messages).

1.2.2 SpamCloud

An important goal of the BGP facet of VIS-SENSE is to confirm, or dismiss, Ramachandran et al. [34]'s suspicions about fly-by spammers. To this end, spam data is provided by Symantec.cloud (formerly known as Message Labs), which sets up and maintains a very large number of spam traps all around the world. All email traffic sent to those spam traps is analyzed by honeypots that extract various features out of the emails, including headers, message content, sender's IP address, name of the bot (if available from CBL [1] rules), embedded URIs, etc. For analysis purposes (e.g. general trends, global spam statistics), the spamtrap traffic is sampled on a daily basis with about 10,000 random samples stored every day in a SQL database. A sampling is required as the actual spam volume intercepted globally and blocked by the company is overwhelming (several billions messages a day) and thus impossible to store entirely. The data collection infrastructure and analytics platform was eventually completely re-engineered to increase the number of email samples being stored and analyzed for intelligence and trend analysis to approximately 4 million spam collected and analyzed on a daily basis after January 2012.

This valuable source of information is leveraged in VIS-SENSE to build a representative spam dataset called SpamCloud, which is automatically fed by the spamtrap data source maintained by *Symantec.cloud.* The SpamCloud data collection process was started in October 2010. Until January 2012, about 10,000 spam samples were automatically copied every day into the data set. Due to the modifications being made to the

Symantec.cloud architecture, the collection process had to be suspended for a few weeks in January 2012. From March 2012, the SpamCloud process was resumed and about 2000 spam samples were inserted every hour, resulting in approximatively 50k new spam messages per day.

For every spam message, a number of spam features are stored within SpamCloud: botnet-related information (if the message was identified as having been sent by a bot), host-related information (IP address, hostname, . . .), and message-related features (from/to domain names, headers, topic, message size, embedded URIs, . . .).

2 BGP Network Analysis and Visual Analytics

This chapters reviews and updates the techniques described in VIS-SENSE deliverables D3.2: Correlation analysis and abnormal event detection module, D4.1 Visual Network Analysis, and D4.2 Visual Correlation Analysis. First, we focus on algorithms developed by the VIS-SENSE consortium in order to detect BGP prefix hijacking. Then, we focus on the specific visualization techniques tailored for these algorithms.

2.1 Network Algorithms

This section contains a recap and an update of the methods developed in D3.2 for BGP analysis.

Techniques for detecting prefix hijacking can be divided into two distinct categories: those based on the *control plane*, i.e. the detection signature depends on information found in a router's routing table and/or messages exchanged with this router; and those based on the *data plane*, i.e. the detection signature is based on the way packets actually flow between an observer and the source. RIPE RIS [5], among others, offers binary dumps of BGP messages exchanged by their routers, as well as snapshots of their routing tables to perform control plane analysis.

Detection techniques from the control plane involve the creation of a model that represents the normal, expected behaviour of a network. Whenever the current view of the network differs from the model, an alert is raised. The complexity and accuracy of the model is then the key element to a good detection scheme.

Detection techniques from the data plane involve active probing of the network topology and/or available live hosts in the monitored network. The core idea is that when a hijacking takes place, significant topology changes should be observed, while the victim network is different from the hijacking network. The way these elements are measured, as well as their diversity, ensures a good detection.

2.1.1 SpamTracer

Data plane measurements can be leveraged to determine the impact of a routing change on the forwarding paths toward a monitored network. We have developed a tool called SpamTracer [48] to monitor the routing behavior of spamming networks by performing traceroute measurements enriched with live BGP information toward networks that have sent spam to Symantec.cloud spam traps. These measurements are performed on a daily basis and repeatedly for a certain period of time after spam is received. We focus on short-lived hijacks by fly-by spammers as observed in [34], i.e. hijacks lasting no longer than one day, thus we set the monitoring period to one week. Currently the system is able to monitor up to ∼8,000 network prefixes everyday with one IP address traced per prefix. By performing measurements on consecutive days for one week for each monitored spam network, data plane paths and BGP routes toward a given network can be compared and analysed in depth to find indications for an ongoing hijack. Because we monitor networks just after spam is received, we expect to observe a routing change as soon as the hijack ends, provided the network was indeed hijacked. The SpamTracer system architecture is depicted in Figure 2.1. The data collection aspects of SpamTracer have been introduced in VIS-SENSE Deliverable 2.2 "Data Collection Infrastructure" and the data analysis aspects have been later described in VIS-SENSE Deliverable 3.2 "Correlation analysis and abnormal event detection module".

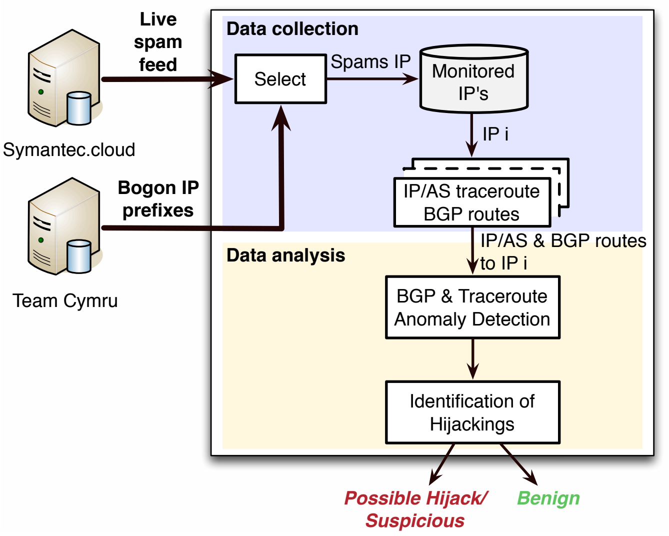

Live spam feed
Symantec.cloud
Bogon IP prefixes
Team Cymru
Data collection
Select
Spams IP
Monitored IP's
IP i
IP/AS traceroute BGP routes
IP/AS & BGP routes to IP i
Data analysis
BGP & Traceroute Anomaly Detection
Identification of Hijackings
Possible Hijack/ Suspicious
Benign

Figure 2.1: SpamTracer system architecture

2.1.2 Spatiotemporal Correlation

Apart from the small scale BGP anomalies that refer to specific prefix or path BGP hijacks, the vast majority of BGP anomalies refer to large scale disturbances and anomalies.

These disturbances affect a vast number of ASes and BGP paths, and have a profound impact upon the Internet operation in terms of both the geographical spread of the phenomena and their duration. Thus, a novel analysis technique based on spatiotemporal correlations between the ASes was introduced in Deliverable 3.2, for the detection and attribution of such phenomena.

More specifically, for each Internet country C_d, taking into account the aggregate set of ASes that provide the interconnection of country C_d with the rest of the world, i.e. all the ISPs that act as the gateways of country C_d, vector V^d is created. This vector is defined as follows:

$$V^d = \begin{bmatrix} NAP\left(AS_1^d\right) \\ \vdots \\ NAP\left(AS_Q^d\right) \end{bmatrix} \tag{2.1}$$

where $[AS_1^d, ..., AS_Q^d]$ is the set of all the ASes that provide connectivity services to country C_d and Q is the size of this set, while $NAP(.)$ is a function that takes as input an AS and returns the number of prefixes that i) are hosted in country C_d and ii) are being announced with this AS as the last hop before any AS hosted in C_d.

Due to the dynamic nature of BGP, the announcement and withdrawal messages change the distribution of the prefixes in the Internet as well as the valid AS-paths for each prefix. Thus, vector V^d is a function of time, in which case a new vector can be defined that captures the dynamic characteristics and evolution of vector V^d as follows:

$$R^d = \begin{bmatrix} V_{t_1}^d & \cdots & V_{t_m}^d \end{bmatrix} = \begin{bmatrix} NAP\left(AS_1^d, t_1\right) & \cdots & NAP\left(AS_1^d, t_m\right) \\ \vdots & \ddots & \vdots \\ NAP\left(AS_Q^d, t_1\right) & \cdots & NAP\left(AS_Q^d, t_m\right) \end{bmatrix} \tag{2.2}$$

Each row of table R^d corresponds to a snapshot of the routing records of the prefixes that concern a specific country C_d. Toward acquiring a quantitative metric for the evaluation of the abnormality of each time period, the Pearson Coefficient (ρ) among all the columns of R^d is calculated. As a result, the Pearson Coefficient matrix of country C_d against time is formulated as follows:

$$\rho^d = \begin{bmatrix} \rho\left(V_{t_1}^d, V_{t_1}^d\right) & \cdots & \rho\left(V_{t_1}^d, V_{t_m}^d\right) \\ \vdots & \ddots & \vdots \\ \rho\left(V_{t_m}^d, V_{t_1}^d\right) & \cdots & \rho\left(V_{t_m}^d, V_{t_m}^d\right) \end{bmatrix} \tag{2.3}$$

The Pearson Coefficient ($-1 \leq \rho \leq 1$) is a metric of the linear dependence between two vectors. Thus, values of this metric ($\rho(V^d_{t_i}, V^d_{t_j})$) much lower than 1, point to large deviation between vectors $V^d_{t_i}$ and $V^d_{t_j}$, and as a result large routing change regarding country C_d between the time instances t_i and t_j.

Thus, matrix ρ^d provides an easy to use analytical method to reveal any time windows with substantial alterations regarding the BGP behavior and the corresponding routing operations that are related to country C_d.

2.1.3 Geospatial Correlation

There are generally two types of BGP hijacks: AS-path hijack and prefix hijack. In order to acquire a quantitative evaluation of the BGP activity and detect possible BGP hijacks a novel method was introduced in Deliverable 3.2 based on the geospatial correlation of the different ASes. The basic notion behind this approach lies within the fact that external Internet routing bears extensive geographic characteristics. Hence, any lawful BGP activity between two ASes should present a geographic coherence.

To begin with, two sets are defined, the set of ASes:

$$A = \{a_i \mid i \in \{1, D_A\}\} \tag{2.4}$$

and the set of Countries:

$$C = \{c_i \mid i \in \{1, D_C\}\} \tag{2.5}$$

where D_A is the number of ASes in the Internet, and D_C the number of country domains in the Internet.

Each Internet AS has a hosting country, i.e. the country where the AS is administratively positioned and the majority of its network infrastructure (e.g. routers) are located.

Taking into account both types of BGP hijack (AS-path hijack and prefix hijack), two distinct sets are defined:

- I^d is the set of all the countries that host all the intermediate ASes that are traversed in order for IP traffic from the monitoring country (country of origin of the monitoring AS) to reach the destination country c_d for every announced prefix.
- U^d is the set of countries that host all the ASes that announced prefixes previously originated by country c_d.

A BGP anomaly is considered to be any activity that exhibits inconsistencies in contradiction to the usual pattern. By exploiting statistics on the two sets, I^d and U^d, any deviation from the usual-normal behavior can be detected, which as a result leads to the detection of BGP anomalies. Under this consideration, two different cases of BGP activities are quantitatively evaluated using the proposed approach: BGP path alteration events and MOAS (Multiple Origin AS) events. Specifically for each case, two sets of metrics are defined based on statistical analysis:

1. BGP path alteration events: For each path alteration event the following metrics are defined.
   - CAP: The probability of appearance of the *Intermediate-Country* $c_i \in I^d$ within the AS-path toward the specific *Origin-Country* c_d.
   - CAPZ: The Z-score of the aforementioned probability.
   - CGL: The geographic deviation introduced by the *Intermediate-Country* $c_i \in I^d$ within the AS-path toward the specific *Origin-Country* c_d. It is defined as the ratio of the aggregate geographic distance between the *Monitoring-Country* and the *Origin-Country* when the route traverses the *Intermediate-Country* against the ideal direct path between the *Monitoring* and the *Origin Countries.*
   - CGLZ: The Z-score of the aforementioned CGL feature.
2. BGP MOAS events: For each MOAS event the following metrics are defined.
   - CAP: The probability of appearance of country $c_j \in U^d$ in a MOAS incident concerning the specific *Origin-Country* c_d.
   - CAPZ: The Z-score of the aforementioned probability.
   - CGL: The geographic deviation introduced by country $c_j \in U^d$ in a MOAS incident toward the specific *Origin-Country* c_d. It is defined as the ratio of the aggregate geographic distance between the *Monitoring-Country* and the *Origin-Country* as if the prefix would traverse through country c_j, against the direct path between the *Monitoring* and the *Origin Countries.*
   - CGLZ: The Z-score of the aforementioned CGL feature.

Taking into account the value of these metrics the analyst can derive useful conclusions regarding the malicious nature of the phenomena under investigation. Specifically, the analyst is interested in investigating events (either MOAS or path alteration) which exhibit low CAP and CAPZ values, while also high CGL and CGLZ score values, i.e.

have low probability of appearance while also induce large geographic deviation. These events are considered anomalous and need further investigation using visualization as well as information from additional sources, in order to derive definitive conclusions regarding their malicious nature.

2.1.4 Geospatial Correlation for the detection of anomalous BGP announcements

The analysis presented in Section 2.1.3 utilizes the geographical position of the ASes in the globe, in order to define features capable of quantifying the degree of anomaly of the BGP path change and MOAS events. Using these features the analyst can focus on the most interesting events for further analysis using additional information sources, or visualization methods developed within the VIS-SENSE framework.

This is very useful, but it only characterizes either path change or MOAS events. In other words, in the absence of these events, a malicious activity is not be visible, and the attacker could evade detection. In order to overcome this drawback, the analysis presented in Section 2.1.3 has been adapted to characterize every BGP announcement regardless of the presence of an event (path change or MOAS).

Using the same notation as in Section 2.1.3, the following features are defined to quantify the degree of anomaly of each BGP announcement with regards to the traversed ASes in the AS-path:

- CAP: The probability of appearance of the *Intermediate-Country* $c_i \in I^d$ within the AS-path toward the specific *Origin-Country* c_d.

- CAPZ: The Z-score of the aforementioned probability.

- CGL: The geographic deviation introduced by the *Intermediate-Country* $c_i \in I^d$ within the AS-path toward the specific *Origin-Country* c_d. It is defined as the ratio of the aggregate geographic distance between the *Monitoring-Country* and the *Origin-Country* when the route traverses the *Intermediate-Country* against the ideal direct path between the *Monitoring* and the *Origin Countries.*

- CGLZ: The Z-score of the aforementioned CGL feature.

It must be noted that within a given day of BGP activity the number of BGP announcements per monitoring point can reach up to 10 million. Thus, it is not efficient to store, analyze and visualize all this information. To this end, only the most suspicious announcements (first 3,000) are stored for further investigation.

2.1.5 Inter-AS relationships

The analysis based on geospatial correlations between ASes operates on per country level and evaluates the BGP activity referring to inter-Country phenomena. In order to over come this limitation and also capture intra-Country activity, another analysis was introduced in Deliverable 3.2 for BGP hijack detection and attribution on the basis of inferred inter-AS relationships.

Despite of the fact that Internet is built utilizing a mesh architecture, the information exchange and transmission is limited by commercial relationships between the ASes. These relationships dictate the existence of an information flow between ASes, as well as its direction. By extensively examining the inter-AS relationships it is possible to quantitatively evaluate the BGP phenomena under investigation. The analysis that follows concerns only MOAS events.

In this context three major types of AS relationship exist in the Internet infrastructure [19]:

1. Provider-To-Customer (p2c) or Customer-To-Provider (c2p). A Provider offers transit connectivity to IP traffic of its Customers
2. Peer-To-Peer (p2p). Two ASes can establish an agreement for mutual exchange of traffic on a quid pro quo basis. The involved peers forward to each other only traffic regarding either themselves or their customers.
3. Sibling-To-Sibling (s2s). Two ASes that administratively belong to the same organization can be connected through a direct link. Two siblings exchange only IP traffic of their own origin.

Utilizing the aforementioned AS relationships, the valid BGP paths are considered to be the paths that are comprised of the following sequence of sub-routes [19]:

- Zero or more c2p links, followed by
- Zero or one p2p link, followed by
- Zero or more p2c links

Upon these definitions, a new metric is defined that captures the business relationships between all the ASes and in particular those involved in MOAS events, so as to allow for the evaluation of the maliciousness of each incident. This metric is defined as the geodesic distance (*GDO*) between two ASes in the AS topology, which is the shortest path that connects the two ASes, given that this path also obeys to the rules of valid BGP paths presented in the previous paragraph.

Thus, based on the GDO metric, a neighborhood of ASes is indirectly defined for each AS, comprised of the ASes that are relatively close (low GDO value). The intuition is that any ordinary and legitimate MOAS event should entail ASes of the same neighborhood. While on the other hand the more distant the ASes are, the more suspicious is a MOAS event between them, which as a result may refer to a potential hijack.

In addition to identifying the relationship of the ASes involved in a MOAS incident, a more comprehensive view would be provided by investigating the relationship of the new owner AS with the AS that originate neighboring prefixes. The term of neighboring prefixes refers to the $2 * k$ nearest prefixes that are announced in BGP and are in close proximity to the prefix under investigation in the IP space. The first k prefixes are higher and the latter k prefixes are lower than the prefix under investigation in the IP space. Since, the neighboring prefixes might belong to many different ASes, there are three possibilities for the calculation of the geodesic distance between them and the new owner ASes: 1) Take the maximum geodesic distance ($ECOPN$), 2) take the minimum geodesic distance ($ECOPN^{min}$), and 3) take the mean vale of all the geodesic distances ($ECOPN^{mean}$). These three metrics define the maximum, minimum, and mean eccentricity of the AS that performed the MOAS incident.

Thus, the eccentricity metrics provide a numerical insight into the relationship of the AS under investigation with the prefix it announces, regardless of its actual relationship with the former owner of this prefix. The higher the eccentricity metrics the lower the relationship of the new owner AS with the announced prefix, and as a consequence the higher the possibility of a malicious action.

To sum up, four metrics are defined for BGP MOAS hijack detection and attribution on the basis of inferred inter-AS relationships:

- GDO: Geodesic Distance between the former and the later origin-AS of the prefix in contention

- $ECOPN$: The maximum eccentricity between the later origin-AS (suspected attacker) and the set of the origin-ASes of the Prefix Neighbors of the prefix in contention

- $ECOPN^{min}$: The minimum eccentricity between the later origin-AS (suspected attacker) and the set of the origin-ASes of the Prefix Neighbors of the prefix in contention

- $ECOPN^{mean}$: The mean eccentricity between the later origin-AS (suspected attacker) and the set of the origin-ASes of the Prefix Neighbors of the prefix in contention

2.1.6 MOAS Filtering

The contents of this section present the MOAS filtering method integrated within BGPDB. This Section first extensively studies the standard routing practices on the global Internet. These findings are then used to filter them out of the suspicious set of suspect prefixes.

2.1.6.1 Network Analysis

MOASes are situations in which a single prefix is originated by multiple ASNs at a given point in time. This situation is usually considered to be the result of prefix multihoming, and goes against RFC1930's [20] recommendation. Here, we study the root causes of MOASes because they reflect standard routing practices. Moreover, most of MOAS occurences are benign [51]. Since our ultimate goal is to model these practices in order to automatically filter out cases that result from them, we also need to look at the evolution of these practices over the years. For this, we rely on the analysis of one full year of measurement collected 10 years apart, in 2002 and 2012 respectively, in order to underline discrepancies that may arise due to changes in standard practices, or due to the global evolution of the Internet.

In 2012, only 5.6% of the announced prefixes had a MOAS at one point. Misconfigurations aside, in case of multihoming or anycasting, one would expect MOASes to be long-lasting. This seems to be confirmed by the mean duration (30.9 days) provided by [51], because MOAS are meant to increase the reachability of a prefix.

To ease interpretation of the observed MOASes, we introduce a taxonomy of MOAS into **provider-customer**, **classical**, and **me-too** MOAS patterns, and study their prevalence and temporal characteristics.

Previous Work on MOAS

Zhao et al. [51] pioneered the analysis of MOASes, and analysed BGP data between late 1997 and mid 2001. This analysis concluded that 36% of the MOASes were one-time and lasted less than a day, 30% of which were attributed to a single misconfiguration event. Excluding those, the average MOAS duration was 30.9 days. For those that lasted over 9 days, the mean duration was 107.5d. These figures are computed from the total duration of a MOAS for a given prefix p, regardless of the origin ASes involved in it, or if the occurrences were continuous or not.

The authors then discuss a number of reasons – other than misconfiguration or maliciousness – for why a prefix would be originated from multiple ASes: prefixes associated with an Internet exchange point (IX) may be advertised by all the ASes within the IX, since they are reachable through all of them. Multihoming without BGP (i.e. via static links or some IGP protocol) also leads to MOAS, since the prefixes are then announced by the upstream providers. Multihoming with BGP, but with a private ASN yields the same result. Anycasting can also lead to MOAS. Finally, since prefix aggregation in BGP transforms the AS path into an AS set (in which the order of AS numbers is random), some artificial MOAS can be observed.

Chin [15] revisited the work of Zhao studying three weeks of data in January 2007, and found an average lifetime of MOASes to be 13.25 hours. Chin then proposed new reasons for why a prefix would be a MOAS: multinational companies may advertise prefixes from various branches in different countries, and such an organization possibly owns multiple AS numbers. Companies may also host their servers in data centers, announcing the prefixes both, from the data center and from their offices. Developing countries that use satellite links and simultaneously use different providers have their prefixes announced by these providers, resulting in a MOAS conflict.

A MOAS can, of course, be the result of a malicious attack against the routing infrastructure. As a result, the problem has been widely discussed in the literature related to prefix hijacking, such as [27, 42]. One of the first systems devoted to the detection and notification of MOAS was PHAS [27]. Latest-generation of tools, such as Argus [42], go far beyond MOAS monitoring in order to detect prefix hijackings. However these hijack papers usually focus on the threat posed by MOASes, not on their behaviour or classification.

Definitions

A **MOAS** (Multiple-Origin AS) is the result of a prefix p being simultaneously originated from multiple ASes. In other words, at a given point in time, the AS paths for p end by a set $\mathcal{O}(p)$ of multiple origin ASNs, so that $\mathcal{O}(p) = \{a_1, \ldots, a_n\}$. For example, using Figure 2.2, $\mathcal{O}_{[t_0,t_1[}(p) = \{1\}$, $\mathcal{O}_{[t_1,t_2[}(p) = \{1, 2\}$, $\mathcal{O}_{]t_2,t_3[}(p) = \{2\}$, $\mathcal{O}_{]t_3,t_4[}(p) = \varnothing$, and so on. It is important to stress that these MOAS situations only occur for the same prefix p. In particular, any prefix q more specific than p with a different origin than that of p is not defined as a MOAS, but as a *sub*-MOAS, which we will not discuss in this document. If a prefix p creates a MOAS situation, we will say that p is a MOAS.

The literature often classifies MOASes according to their duration with the following

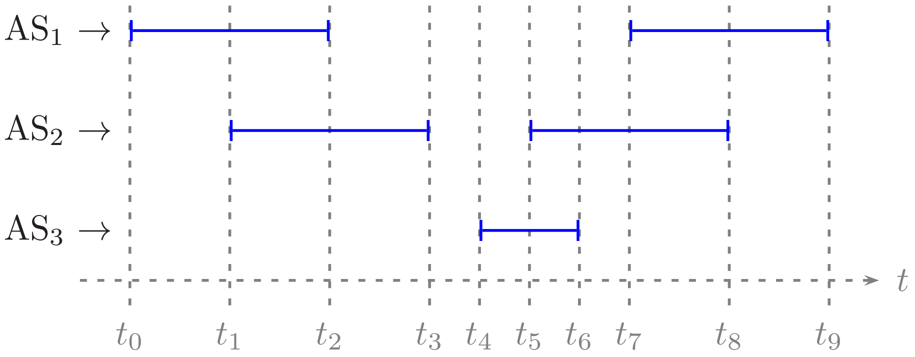


Figure 2.2: Example of announcements for a prefix p

terminology: **short-lived MOASes** last less than 24 hours, **long-lived MOASes** last more than 1 day. We call **very short-lived MOASes** occurrences whose duration are 6 minutes or less; and **very long-lived MOASes** occurrences whose duration are longer than a month.

If p is not a MOAS, but p is still present in the routing tables, p is a **SOAS** (Single-Origin AS), meaning that p is originated by a single AS. In Figure 2.2, this happens during $[t_0, t_1[$, for example. If p is not included in the routing tables, we will say that p is **down** (Figure 2.2 during $]t_3, t_4[$). Please note that p being down does not necessarily imply that traffic destined to p will result in no-route-to-host errors, because a covering less specific prefix of p could be used to forward the traffic. By contrast, a prefix is **up** whenever it is a SOAS or a MOAS.

We define the **lifetime** of a prefix p as the difference between the timestamp at which the prefix was last withdrawn (that is, the timestamp at which the prefix goes down for the last time) and the timestamp at which the prefix was first announced(i.e. the first time it went up during our observation). The lifetime of p in Figure 2.2 is simply $t_9 - t_0$. On the other hand, the **uptime** of p is defined as the total duration during which the prefix was up. In Figure 2.2, the uptime of p is $t_3 - t_0 + t_9 - t_4$.

BGP Dataset

In order to study MOASes, we use data from RIPE RIS [5]; more specifically we use the route collector located in Amsterdam (`rrc00`). We retrieve the update messages, and simulate BGP's operations. More precisely, we maintain a routing table for each peer – similar to BGP's Adj-RIB-In – the *adjacent routing table*. Each route announced by a peer is added to that peer's adjacent routing table. Whenever a withdrawal is received for a prefix, every route to that prefix is removed from the peer's adjacent routing table.

Since we are not interested in routing traffic, we do not try to select the best route among all the existing ones. We are, however, interested in knowing if a prefix p is up, i.e. if p is present in any of the adjacent routing tables.

The set of origins $\mathcal{O}(p)$ associated with prefix p is composed of the union of all the origins included in all of the AS paths of each adjacent routing table. If the cardinality of $\mathcal{O}(p)$ is larger than 1, there is a MOAS for p. For example, in Figure 2.2, during $[t_0, t_1[$, $\mathcal{O}_{[t_0,t_1[} = \{1\}$ whose cardinality is 1, and the prefix is a SOAS. During $[t_1, t_2[$, $\mathcal{O}_{[t_1,t_2[}(p) = \{1, 2\}$ whose cardinality is 2, and the prefix is a MOAS. Finally, during $]t_3, t_4[$, $\mathcal{O}_{]t_3,t_4[}(p) = \varnothing$ whose cardinality is 0, and the prefix is down.

General Results

	MOAS prefix					
	uptime			**lifetime**		
	$\boldsymbol{\mu}$	**CoV**	$\boldsymbol{q_{50}}$	$\boldsymbol{\mu}$	**CoV**	$\boldsymbol{q_{50}}$
2002	328d	0.25	363d	334d	0.23	364d
2012	308d	0.34	364d	317d	0.31	364d

	SOAS prefix					
	uptime			**lifetime**		
	$\boldsymbol{\mu}$	**CoV**	$\boldsymbol{q_{50}}$	$\boldsymbol{\mu}$	**CoV**	$\boldsymbol{q_{50}}$
2002	146d	1.11	37d	172d	0.89	146d
2012	223d	0.72	348d	239d	0.65	360d

Table 2.1: General statistics on BGP data for 2002 and 2012

During the year of 2002, almost 310k different prefixes were announced, less than 9% of which presented (at least) one MOAS. In 2012, there were almost 765k distinct announced prefixes, less than 6% of which presented (at least) one MOAS. These figures suggest that, while both, the number of global prefixes and the number of MOAS prefixes increased in 10 years, their proportion has decreased.

Table 2.1 shows the mean (μ), coefficient of variation (CoV), and median (q_{50}) durations for the uptime and the lifetime of both MOAS and SOAS prefixes during 2002 and 2012. The coefficient of variation is defined as the standard deviation divided by the mean. If the CoV value is lower than 1, the variable is considered to have a *low* variance. Mean values for MOAS prefixes in both, 2002 and 2012 are significantly higher than the values for SOAS prefix in terms of uptime and lifetime. This justifies the use of MOAS to improve the reachability of a prefix. In particular, median uptime and lifetime of MOAS prefixes are both close to 1 year, meaning that 50% of those prefixes were seen over the entire observation period. The mean and median value for SOAS prefixes in 2012 – both close to 1 year – are also much higher than those in 2002, where the median uptime of 37d is very low compared to the observation period of 1 year, and to a median lifetime of 146d. These (low) figures for 2002 are in line with the ones presented in [44].

		μ	**CoV**	**q_{50}**
MOAS (per event)	**2002**	33d	2.23	22h
	2012	48d	1.88	26h
Short-lived MOAS (per event)	**2002**	133mn	2.26	9.3mn
	2012	101mn	2.60	3.13mn
MOAS (per prefix)	**2002**	111d	1.01	71d
	2012	125d	0.95	90d
Provider-Customer MOAS	**2002**	36d	2.16	22h
	2012	48d	1.89	7h
Classical MOAS	**2002**	26d	2.41	22h
	2012	48d	1.83	7d
Me-Too MOAS	**2002**	170d	0.83	148d
	2012	126d	0.95	106d

Table 2.2: Duration of MOAS events

MOAS per Event

In this section, we consider MOAS events on their own, as a set of distinct independant events, independent of the prefix to which they are associated. For example, we consider independently the 3 MOAS depicted in Figure 2.2 during $]t_1, t_2[$, $]t_5, t_6[$, and $]t_7, t_8[$. The **MOAS duration** (per event) is the duration of a single event. In Figure 2.2, the durations of the three MOAS events are $t_2 - t_1$, $t_6 - t_5$, and $t_8 - t_7$.

Figure 2.3 depicts the duration of MOAS events in 2002 and in 2012. Almost 50% of the MOASes were short-lived; very short-lived events appear to be more prevalent in 2012 with 11% of MOASes that last 1 second. MOAS duration information for 2002 and 2012 are available in the first row of Table 2.2. The large difference between the mean and the median shows how prevalent short-duration events are. The equivalent values for short-lived MOASes are given in the next row.

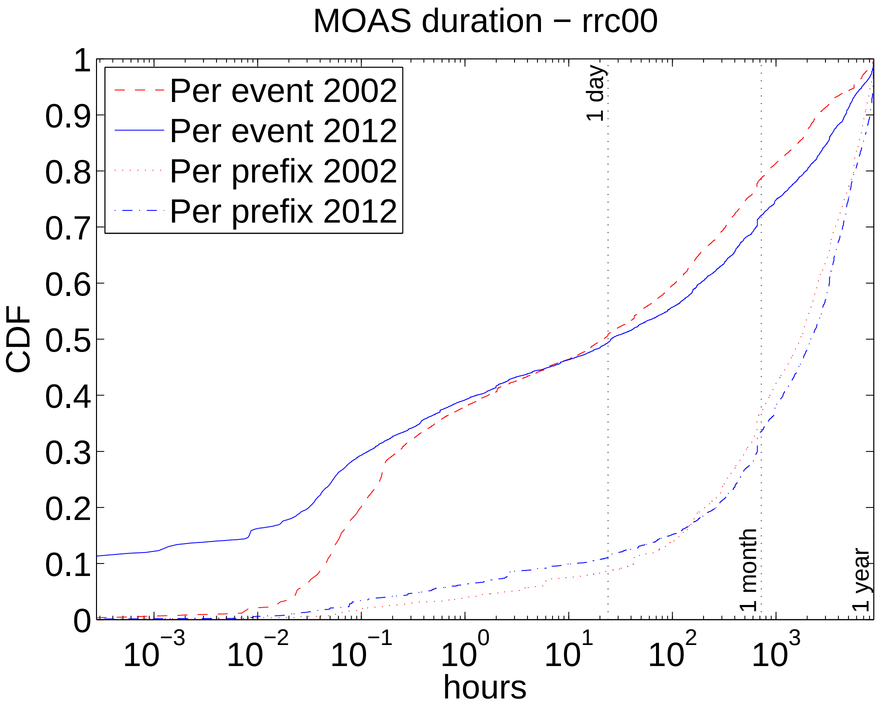


Figure 2.3: MOAS duration

MOAS per Prefix

In this section, we consider MOASes grouped by the prefix for which they appeared. Different MOAS situations may appear over the course of the observation period for a single prefix p. We say two MOASes associated with a prefix p are **distinct** if the origin sets $\mathcal{O}(p)$ are different for the two MOASes. For example, in Figure 2.2, prefix p has 3 MOASes: during $[t_1, t_2[$, $[t_5, t_6[$, and $[t_7, t_8[$. Moreover, $\mathcal{O}_{[t_1,t_2[}(p) = \mathcal{O}_{[t_7,t_8[}(p) \neq \mathcal{O}_{[t_5,t_6[}(p)$. So, even though Figure 2.2 depicts 3 MOASes, only 2 of them are distinct in the sense that they involve different ASes. Furthermore, the duration of **MOASes per prefix** is the sum of the durations of the individual MOAS events associated with this prefix. Using Figure 2.2, the MOAS duration for prefix p is $t_2 - t_1 + t_6 - t_5 + t_8 - t_7$. In the remainder of this section, unless explicitly stated, duration means the duration of the MOAS *per prefix*.

Figure 2.3 plots the duration of MOASes per prefix. Only around 10% of the MOASes are short-lived, which heavily contrasts with the 50% previously presented when considering each MOAS event on its own. This implies that certain prefixes must have many MOAS events. On the other hand, if a MOAS event is due to a misconfiguration, one should not see multiple recurrent MOAS events for the same prefix. This is further confirmed by Figures 2.5 and 2.5, where the number of MOASes and the number of distinct MOASes per prefix are plotted for 2002 and 2012 respectively. The prefixes are sorted by decreasing number of MOAS events. For the first 1000 prefixes with the most MOAS events, the mean and median duration of single MOAS events is very small (in the order of a few minutes or less).

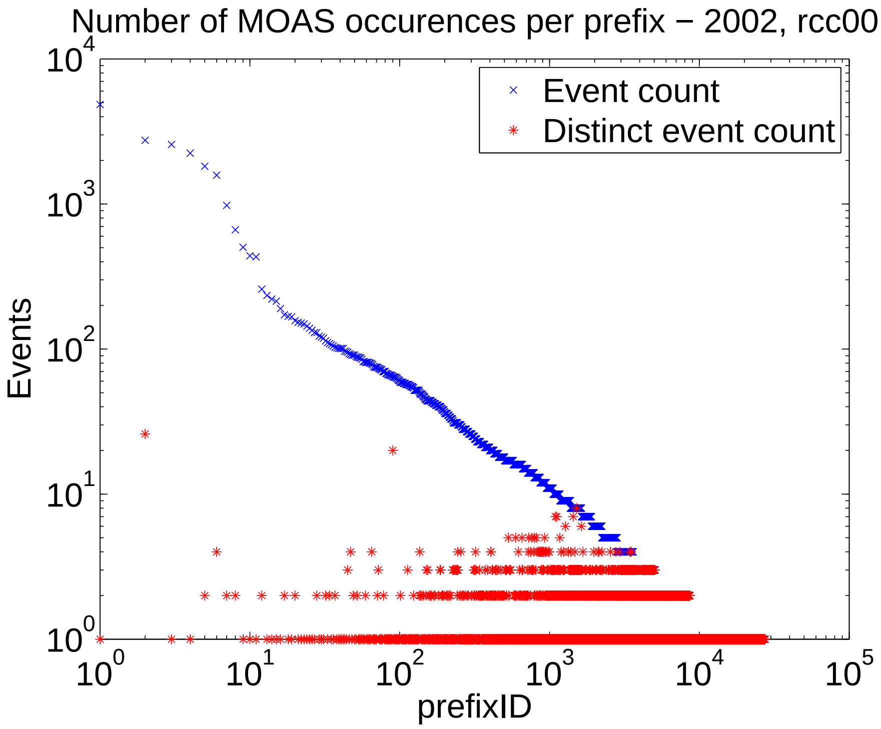


Figure 2.4: Number of MOAS events per prefix

Another implication of Figure 2.3 is that short-lived MOASes are not necessarily the result of misconfigurations, since only the sum of numerous small events concerning the same prefix can raise the MOAS duration per prefix as much. If we accept that misconfigurations are the result of operator errors, the prefixes affected by the misconfigurations should be random. Misconfigurations are expected to be short-lived and "one shot"

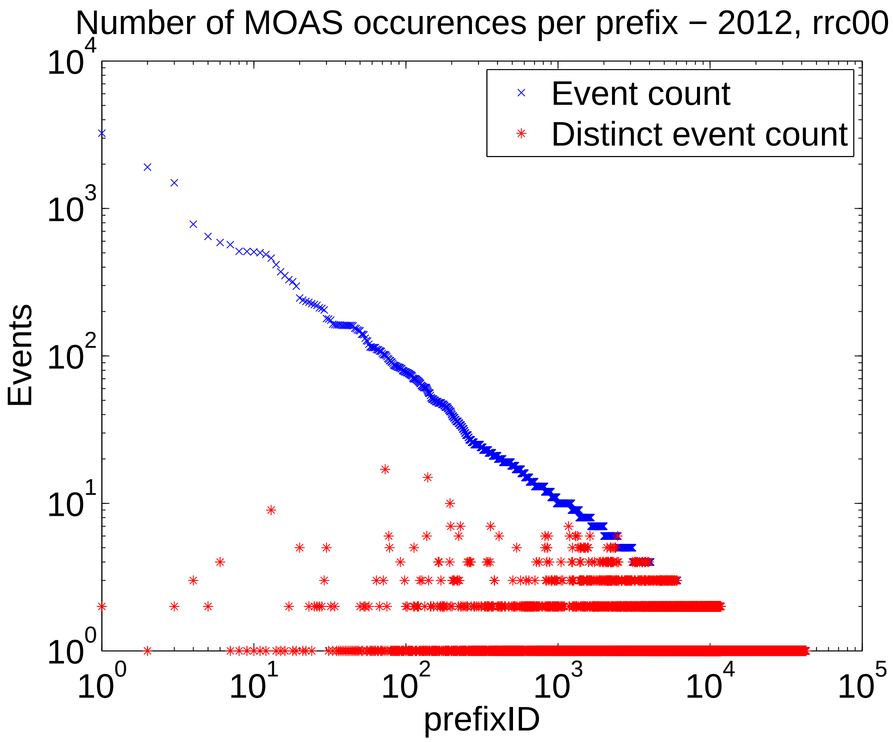


Figure 2.5: Number of MOAS events per prefix

because the operator usually uncovers their mistake right after the new configuration has been saved. If most short-lived MOAS events were the result of misconfigurations, grouping these events by prefix would not modify the distribution as much as shown in Figure 2.3, because there would be more different prefixes affected. As a result, the curves in Figure 2.3 would not show such a drastic difference.

Figure 2.5 shows that, for approximately 1000 prefixes out of the 43k MOAS prefixes, the number of *distinct* MOASes is significantly lower than the number of MOAS events. Some of these prefixes only have 1 distinct MOAS. In this case, there is a continuous flipping between SOAS and MOAS announcements. This can be either the result of one of the origin AS being unstable, or of a BGP router – located between the origin network and the collector – that continuously keeps on changing its mind on the best route. We found one example where a peer of `rrc00` kept flooding `rrc00` with around 2000 update messages for one prefix within a couple of days, while the other peers remained unanimously quiet about it. By looking at the AS paths of these messages, we saw an oscillation between the two origins. Since only that one peer was doing this, we suspect the BGP message was not triggered by the origin networks. The middle section of the AS paths showed that all of these messages went through the next-hop to destination of our peer. We suspect that, due to a loosely-defined routing policy, which results in the two routes having a similar preference, that peer's next-hop kept on changing its mind about which path to destination should be taken. Depending on the routes forwarded by our other peers, it could have been that this flooding continuously in a SOAS/MOAS cycling completely independent of the origins' stabilities.

The mean and median MOAS durations per prefix in 2002 and 2012 are detailed in Table 2.2. We clearly see that both the mean and median values for the MOAS per prefix are a lot larger than individual MOAS event durations. This is, once again, the result of the aggregation of the many short-lived events per prefix.

We also considered the fraction of time in MOAS state for a prefix was over the total uptime of the prefix. One might expect MOAS prefixes to remain in MOAS state during most of their uptime in order to maximize reachability, particularly in case of path failure. However, the distribution of the fraction of time in MOAS state distribution is uniform, which contradicts this expectation. This can be explained by the use of *transient* MOAS configurations. By doing a temporal analysis of the topological evolution of MOAS networks, we witnessed multiple cases of stub networks switching their upstreams AS provider. Normally, this operation happens as follows. Originally, prefix p is announced by ISP A. At some point in time, the owners of p find it more advantageous to use ISP B. In order to avoid any service disruption, p remains connected to (and announced by) A while things are being set up with B (i.e. connecting p to B), and then also starting announcing it from B. This results in a MOAS. After some time (days, weeks), p is

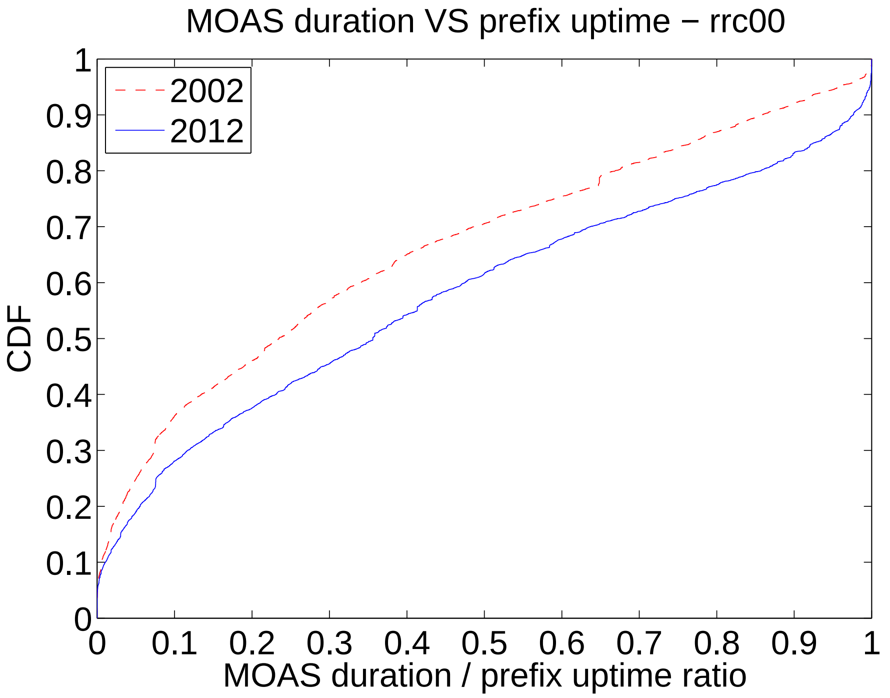


Figure 2.6: Total MOAS duration for the prefix VS prefix uptime

disconnected from A, and remains exclusively announced and reachable through B. A similar situation was encountered where a set of prefixes p_i belonging to different entities all evolved in the same manner, with the same set of transient MOAS conflicts. The reason was a topology change at the ISP used by the prefix owners. In this situation, the prefixes were connected via multiple origins ASes for a couple of weeks, and then a part of the topology graph was pruned, leaving the prefixes effectively reachable via a single AS. The difference between these two situations is that, in the second case, the prefix owners are not the ones who decide to change the way they are connected to the Internet. Moreover, the number of prefixes involved is much higher, since it involves all the prefixes hosted by the service provider. The actual duration of this type of scenario depends on *human actions*, not on technical constraints. We believe this is the major reason between the uniformness of the distribution of the fraction of time in MOAS state.

2.1.6.2 MOAS Patterns

By analysing the AS-level graph of a prefix with MOASes, we were able to extract a set of patterns that result from MOAS announcements. This led us to a topology of MOAS that we present now.

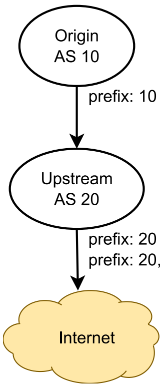


Figure 2.7: Provider-Customer MOAS pattern

The first pattern, depicted in Figure 2.7, shows a situation where both, the prefix owner and its upstream are announcing the prefix. We call this MOAS a **Provider-Customer MOAS**. Even though figure 2.7 only depicts one upstream, we saw cases where upstreams of the upstream were also announcing that prefix. The mean and

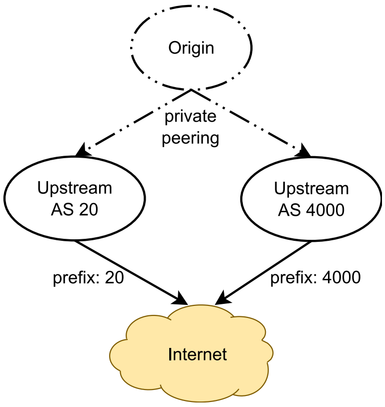


Figure 2.8: Classical MOAS pattern

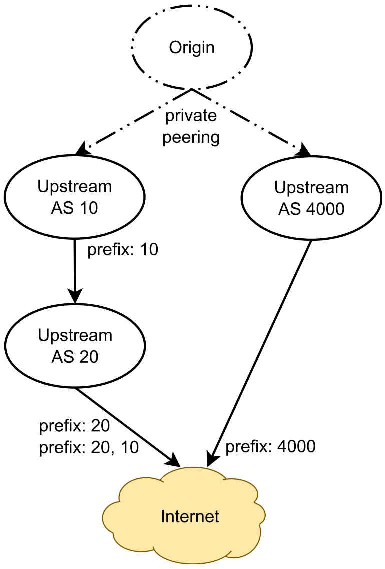


Figure 2.9: Me-Too MOAS pattern

median durations for provider-customer-caused MOASes are presented in Table 2.2. The median durations – 21.5h in 2002 and 7.4h in 2012 – suggest that a non-negligible part of these MOASes are short-lived. Figure 2.10 shows the distribution of the durations of provider-customer-caused MOASes for 2002 and 2012. In both cases, between 55% and 60% of the MOASes are short-lived, but there appears to be more very short-lived events in 2012.

Table 2.3 shows the proportion of provider-customer-caused MOASes among all MOASes. It is surprisingly high with around 70%, even though there appears to be no practical use for such an announcement since the traffic necessarily has to cross the upstream's network before reaching destination. In a sense, we can consider those MOASes as fake MOASes, because there is only one physical path leading to the prefix.

In order to understand the reasons behind this pattern, we proceeded to a temporal analysis of these events. The best explanation for such short-lived MOASes is a topology change. Originally, the prefix is first announced by the upstream, but assigned to the customer (e.g. Figure 2.2, during $[t_0, t_1[$). At some point, the customer decides to handle routing on its own, and acquires its own AS number and starts BGP peering with the upstream. At this point, there is a MOAS (Figure 2.2, during $[t_1, t_2[$). Eventually, the upstream withdraws the announcement of the prefix, leaving only the owner's announcement in the routing tables (Figure 2.2, during $[t_2, t_3]$). In this case, the MOAS was the side-effect of a real topology change. We suspect that long-standing MOASes are due to untouched configurations, i.e. the old adage "if it ain't broke, don't fix it", or that the upstream's policy is to announce all of its customers' prefixes. Another possibility could be that the upstream is hosting some fail-over services on behalf of its customer, like web servers, although we have no evidence of such setups.

The second pattern, depicted in Figure 2.8, is often presented as the **classical MOAS** pattern, as there are two distinct AS paths leading to the prefix. The mean duration of these MOASes are shown in the penultimate row of Table 2.2. These values suggest that the classical MOASes are longer-lived in 2012 than in 2002. This is confirmed by Figure 2.10 which plots the durations of these events. In 2002, around 50% of them were short-lived, which then decreased to around 35% for 2012.

Table 2.3 shows the proportion of classical MOASes among all MOASes, which is around 25%.

The last pattern, depicted in Figure 2.9, is named **Me-Too MOAS** to underline its "being over-announced" property. It is composed of both of the previous patterns at a single time. We first stumbled on this configuration as a transient configuration, when a prefix owner decided to change upstreams. Originally, the owner's prefix was announced by a tier-1 ISP who used multiple AS numbers: one for its global activities (Figure 2.9, AS20), and one for its local activities (Figure 2.9, AS10). However the ISP

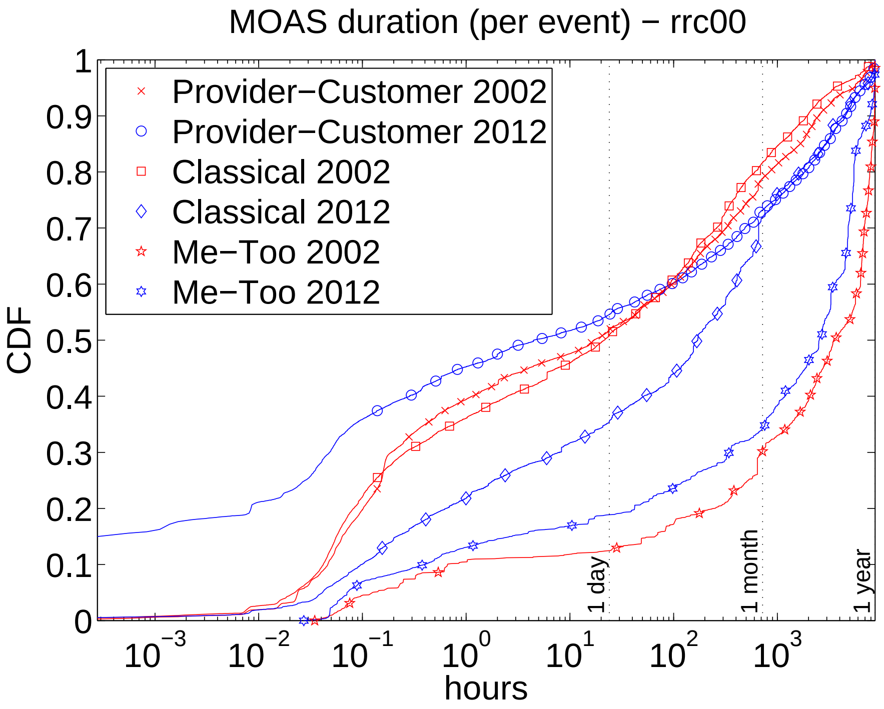


Figure 2.10: MOAS pattern duration per event

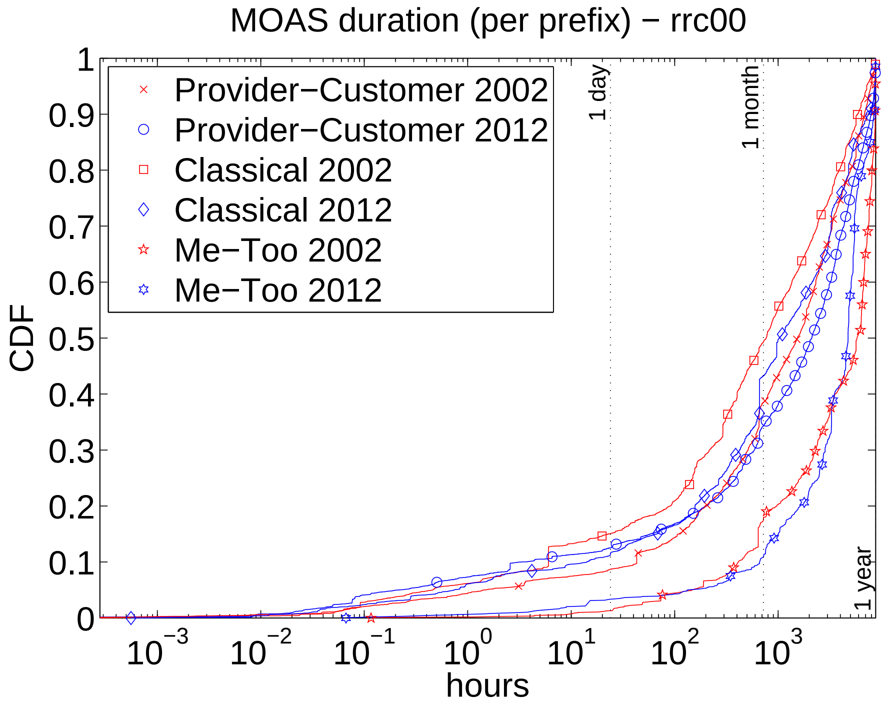


Figure 2.11: MOAS pattern duration per prefix

used both of those AS numbers to originate the prefix, although it needs to go through the local AS from the backbone to reach the customer (this corresponds to a provider-customer-caused MOAS). Then, the user (Figure 2.9, origin) decides to switch their ISP service to another tier-3 ISP (Figure 2.9, AS4000). During the transition, which can last several weeks, the prefix was announced by both the old tier-1 (Figure 2.9, AS10 and AS20) and the new local tier-3 ISP (Figure 2.9, AS4000). This situation, then, presents a provider-customer-caused MOAS with a classical MOAS. This peculiar configuration can be due to a combination of subletting of IP space. Using Figure 2.9 as illustration, the prefix block p is owned by AS20 and AS10 rents it. The whois record associated with p clearly stated that prefix p was part of non-transferable IP addresses. So, because AS20 is the owner, it keeps on announcing p. However, since AS10 rents it, it also announces the prefix. This results in a provider-customer-caused MOAS, i.e. the left-hand side of Figure 2.9. Additionally, AS10 assigned p to one of their customer for use. At some point, this customer chooses to do multihoming and uses AS4000 for that purpose. In return, AS4000 announces p as well, i.e. the right-hand side of Figure 2.9 . It is not clear why the end-customer would intentionally use such non-transferable IP space, or why AS10 would deem it preferable to rent these addresses instead of obtaining a dedicated block. A possible explanation is that the customer started with a small network, which eventually grew bigger. In order to avoid reconfiguring the whole network, the customer was willing to keep on using the same IP block.

The mean and median durations of this pattern is shown in the last row of Table 2.2. These values suggest that me-too MOASes are indeed stable. Figure 2.10 confirms that few of these events are short-lived (around 20% in both cases), and over 60% of them last longer than two months. This can be explained by the fact that this configuration is unlikely to arise from erroneous situations, unlike the previous two patterns since it requires (at least) 3 origin ASes for a single prefix, *with* a provider-customer relation among two of them.

Table 2.3 shows the proportion of events for each MOAS pattern for 2002 and 2012. More than two thirds of MOAS are of type provider-customer, which means that more than two thirds of MOASes are of questionable use, and the prefixes associated with these events would probably be better off without the MOAS announcement. As previously conferred from Figure 2.10, a lot of these events are short-lived, and can be attributed to either an AS's policy or a topology change. This observation clarifies the numerous short-lived events unveiled by [51, 15] and often attributed to the result of misconfigurations by the lack of further analysis. On the other hand, me-too MOAS are very rare. If we look at the MOAS pattern not per event but per prefix, the proportions of the different patterns are almost the same, suggesting that only a few prefixes actually exhibit more than a single MOAS pattern.

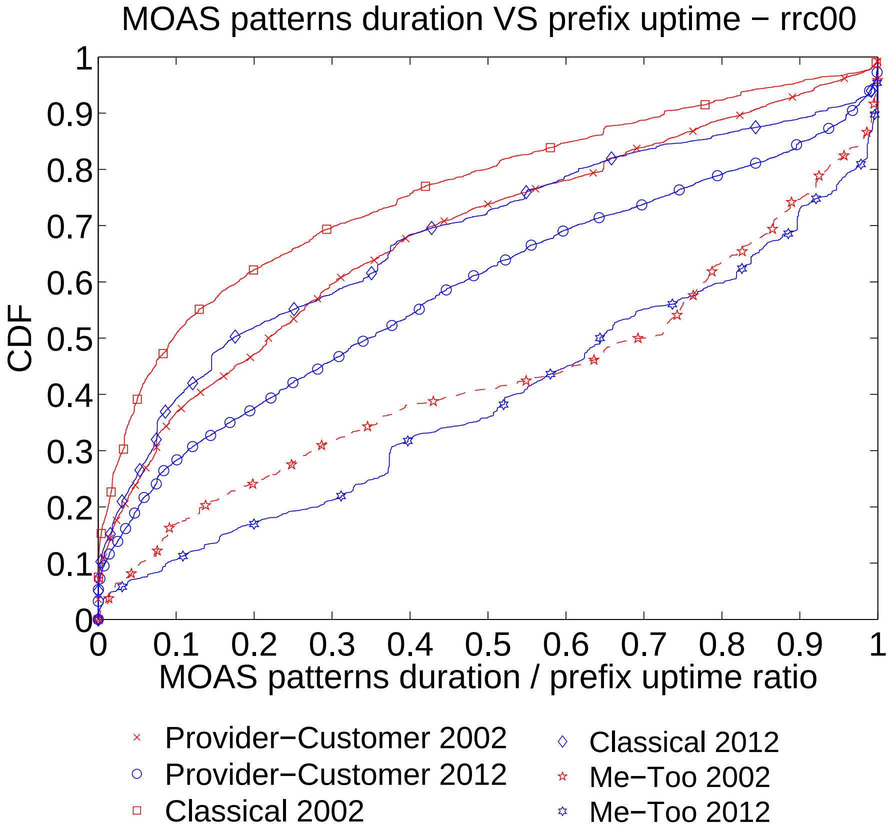


Figure 2.12: Total MOAS pattern duration for the prefix VS prefix uptime

MOAS Pattern Duration per Prefix

Figure 2.11 shows the duration per prefix of the MOAS patterns. Again, the number of short-lived MOASes per prefix is drastically reduced compared to the MOAS durations per event (Figure 2.10). The drastic reduction – from around 50% to less than 15% – for provider-customer MOAS suggests that the conflicts are long-lived, and that the configuration remains unchanged, even though its benefits are questionable. It might just be that some providers always announce their customer's prefixes as a policy, maybe because not every customer uses BGP. Short-lived classical MOAS durations are drastically reduced from around 40% to around 15% when considering MOASes per prefix, suggesting that many short-lived MOAS are triggered by route instabilities and not by misconfigurations.

Per event	**2002**	**2012**
Provider-Customer MOAS	69.21%	74.79%
Classical MOAS	28.84%	23.08%
Me-Too MOAS	1.95%	2.12%

Per prefix	**2002**	**2012**
Provider-Customer MOAS (a)	72.55%	72.63%
Classical MOAS (b)	31.09%	28.6%
Me-Too MOAS (c)	5.5%	3.84%
(a) & (b)	6.37%	3.24%
(a) & (c)	1.95%	1.59%
(b) & (c)	1.37%	0.49%
(a) & (b) & (c)	0.52%	0.24%

Table 2.3: Proportion of occurrences of MOAS patterns

Conclusion

During the last ten years, little has changed in terms of MOAS engineering practices, which appears to be motivated by the "do not change if it works" principle. For instance, 70% of occurring MOASes are of type provider-customer. Their benefit is highly questionable, and because there is only one AS path to reach such a prefix, it does not improve reachability. Classical MOASes are typically used for network engineering reasons, such as multihoming, load balancing, or anycasting. They account for around 30%

of MOASes, and offer multiple AS paths to a prefix. Me-too MOASes are very rare, and are the result of exotic, and possibly out-dated (but still running) router configurations.

In addition, the uniformity of MOAS duration distributions indicates the absence of globally applied guidelines for BGP engineering. Consequently, network operators rely primarily on their experience. This results in a high variation of MOAS duration, because they are the result of human actions, and not of protocols or technical requirements.

By looking at MOASes grouped per prefix, we are able to show that short-lived MOASes are not necessarily the result of router misconfigurations. By comparing the number of MOAS events per prefix with the number of *distinct* MOAS events per prefix, we conclude short-lived MOASes are mostly triggered by route instabilities. The remaining short-lived MOASes are likely due to misconfigurations, though. However these events are not as prevalent as other studies suggested: we observe less than 10% of them.

2.1.6.3 Applying to Security

In this paragraph, we use the data analysis presented in the previous Section, and use it in order to build a set of filters on MOAS alerts in order to reduce the set of MOAS events that need to be manually inspected. Please note that this method is presented as a separate module from Geospatial Correlation (Section 2.1.5) and Inter-AS Relationships (Section 2.1.5).

The three MOAS patterns presented in the previous Section – Provider-Customer MOAS, Classical MOAS, and Me-Too MOAS – can each potentially lead to different security implications.

A Provider-Customer MOAS conflict presents absolutely no security threat because there is no reason for which a provider needs to hijack the traffic of one of its customers. This is because the provider is always on the customer's way to/from the Internet. Consequently, should a provider decide to eavesdrop on or temper with the traffic of one of its customers, it can do so without BGP means. As a result, we can disregard any MOAS alert that is raised due to a provider-customer relationship in the Internet. Methods to infer the relationship between two ASes have been presented before, most notably by [19], and are even available as pre-computed input files from organizations such as CAIDA [2]. These methods inferring AS relationships usually output a directed relation (i.e. the provider is assigned as one of the AS, the customer as the other). We are just interested in the existence of such a relationship, not in its direction. By doing this, we avoid raising alerts due to the random order in which we detect the announcing ASes.

Classical MOAS and Me-Too MOAS, on the other hand, cannot be filtered out by their AS-level network topology. Both of these patterns can be the result of a legitimate announcement where the prefix owner's network is connected to its upstreams via a private BGP peering, or some static route configuration. But they could also be the result of the injection of erroneous data in BGP. However, Figures 2.10 and 2.11 show that while short-lived MOAS conflicts are quite numerous (around 50% of event for classical MOAS, and 20% for Me-Too MOAS), the number of short-lived conflicts per prefix is significantly lower (15% for classical MOAS, and less than 5% for Me-Too MOAS). This suggests that the bigger bulk of MOAS events is indeed composed of stable, long-lived prefix-wise MOAS conflicts. In order to filter these out, we rely on the duration of the announcement, based on Pretty-Good BGP's (PGBDP) suggestions [23]. Pretty-Good BGP [23] suggests that the global routing infrastructure's robustness would largely benefit from quarantining new, unknown routes 24h before allowing them to be used to forward IP traffic. This is due to the fact that the network owner has the time to take actions against long-lived erroneous announcements. For example, they can contact the attacker and inform them that they are misbehaving. If this does not lead to a withdrawal of the bad announcement, they can contact the upstream of the attacker in order to try and get them to filter out this announcement. In a similar way, we only start trusting an origin when its uptime for a given prefix is bigger than a threshold T. However, because the topology of the Internet is constantly evolving, we also need to stop trusting origins that have not announced a given prefix for a long time. As a result, we discard any origin AS for a given prefix, regardless of its uptime, if that origin has not announced the prefix in the last 30 days. Using this method, the model for a prefix – containing the trusted origin ASNs along with their uptime – is updated as time goes by, and always sticks to what is currently seen in the network.

The goal of these filters is to take advantage of standard routing practices in order to filter out as many benign MOAS conflicts cases as possible. Compared to well-established methods derived from the well-known PHAS algorithm [27], we suppress between 65% and 80% of alerted prefixes. The remaining cases, composed exclusively of Classical MOAS conflicts (or mixed Me-Too MOAS conflicts), need to be investigated through some other means. This can be through active probing of the alerted network (e.g. using traceroutes), which would underline the existence of two different end-networks if the prefix is indeed under attack.

As a summary, we give here the steps of the MOAS filtering algorithm:

1. Process BGP update messages according to RFC4271 [35]
    - An adjacent routing table is created for each peer of our vantage point(s)
    - Each route to a prefix p announced by a peer N is added to the adjacent routing table associated to N.

- When a prefix p is withdrawn by a peer N, the route to p in the adjacent routing table associated to N are erased.

2. When a prefix p is updated
   - If p is a new prefix (i.e. if p's data is not currently in memory)
     - Fetch the previous origins for p in BGPDB that were seen during the previous 30 days, and that have an uptime longer than 48 hours
     - If no such origin AS exist, the prefix is in *learning mode* for 24 hours, and any incoming origin for p will be accepted as ground-truth. This is to avoid MOAS alerts on prefixes that were not previously announced and that are now affected, or reaffected.
   - Check if the incoming origin for p is different than the one(s) we already know.
     - If it is not, everything is fine.
     - If the origins are different, raise a MOAS alert. MOAS alerts will be raised for that origin until it reaches an uptime longer than 48 hours.
3. When a prefix p is withdrawn
   - If p has been withdrawn by every peer (i.e. if p does not appear in any of the adjacent routing tables), then p is not part of BGP routing tables anymore. p is unannounced, has gone offline. Delete the model associated to p from memory. (It can be recreated from BGPDB later, if needed.)

2.2 Visualizations

The following section describes the main visualizations related to BGP analysis introduced in Work Package 4, which can be used for visual exploration of BGP data and the analysis of analytical results provided by the aforementioned network algorithms.

2.2.1 VisTracer

As described in Deliverable 4.2, *VisTracer* is a visual analytics tool, which combines different visualization techniques with the continuously growing *SpamTracer* database. 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 [43]. 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 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 can then be used for further improving the underlying anomaly detection algorithms.

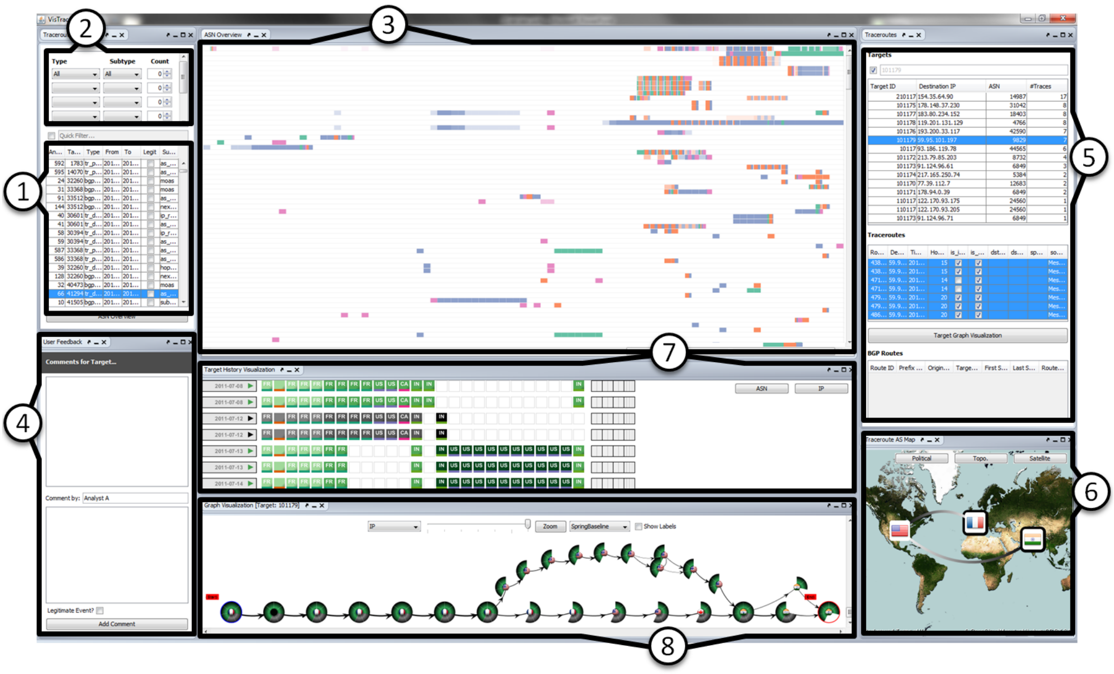


Figure 2.13: Graphical user interface of the *VisTracer* visual analytics tool.

The overall graphical user interface is shown in Figure 2.13. 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.

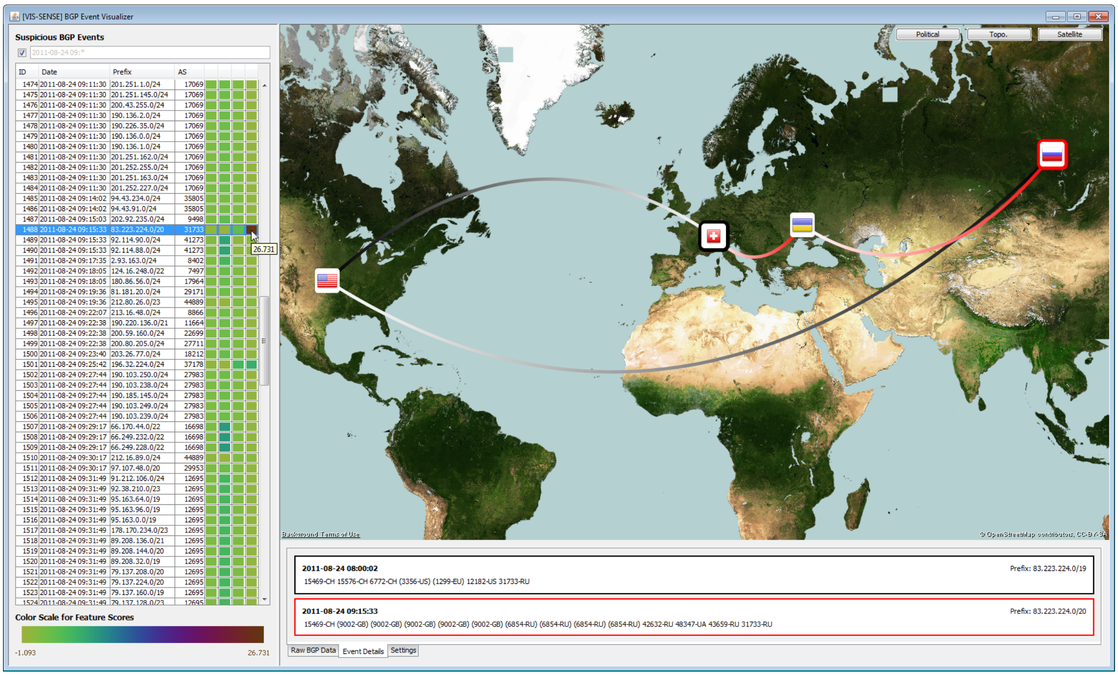
2.2.2 BGP Event Visualizer

BGP Event Visualizer is a tool, which combines strong statistical methods with multiple views to best support an explorative analysis. Of course, detecting deviation from normal behavior can be done automatically. But having a visualization approach helps to understand the suspicious event better and allows an exploratory analysis of the anomaly.

The Statistical Metrics are calculated in the network analytics layer and are extensively described in Deliverable 3.2. The results are then displayed with two different kinds of visualizations. The overview visualization consists of a sortable table view with suspicious BGP events. Each row represents one event with different information dimensions. The columns provide information about the date, the affected prefix and target AS and additionally colored pixels to represent the results of the previously done automatic analysis. The color scale of the pixels represents the locally normalized values of the different metrics like e.g. the Z-Score for a target AS. By visualizing different metrics in the same way, patterns can be easily recognized as a combination of multiple colored pixels.

After detecting an interesting suspicious event this entry can be selected to reveal further details about the underlying connections. A second display uses a graph layout on top of a geographic map to visualize the routing information in a convenient way (Figure 2.14). ASes are represented by small country flags according to their country code information. Curved lines between these ASes are plotted to provide information about the connections. The direction of each line is encoded by the color saturation of the lines reaching from white (no saturation) to black or red respectively.

Visualizing raw BGP update messages is nearly impossible due to the massive amount of data. Statistical preprocessing is, therefore, mandatory to make sense out of the data. However, having the statistical metrics displayed in a table is not sufficient to really make sense out of the data. To detect suspicious BGP update anomalies it is important to provide the security expert with a tool, which tightly couples statistical information with expressive visualizations.



Suspicious BGP Events

☑ 2011-08-24 09:*

ID	Date	Prefix	AS				
1474	2011-08-24 09:11:30	201.251.1.0/24	17069	🟩	🟩	🟩	🟩
1475	2011-08-24 09:11:30	201.251.145.0/24	17069	🟩	🟩	🟩	🟩
1476	2011-08-24 09:11:30	200.43.255.0/24	17069	🟩	🟩	🟩	🟩
1477	2011-08-24 09:11:30	190.136.2.0/24	17069	🟩	🟩	🟩	🟩
1478	2011-08-24 09:11:30	190.226.35.0/24	17069	🟩	🟩	🟩	🟩
1479	2011-08-24 09:11:30	190.136.0.0/24	17069	🟩	🟩	🟩	🟩
1480	2011-08-24 09:11:30	190.136.1.0/24	17069	🟩	🟩	🟩	🟩
1481	2011-08-24 09:11:30	201.251.162.0/24	17069	🟩	🟩	🟩	🟩
1482	2011-08-24 09:11:30	201.252.255.0/24	17069	🟩	🟩	🟩	🟩
1483	2011-08-24 09:11:30	201.251.163.0/24	17069	🟩	🟩	🟩	🟩
1484	2011-08-24 09:11:30	201.252.227.0/24	17069	🟩	🟩	🟩	🟩
1485	2011-08-24 09:14:02	94.43.234.0/24	35805	🟩	🟩	🟩	🟩
1486	2011-08-24 09:14:02	94.43.91.0/24	35805	🟩	🟩	🟩	🟩
1487	2011-08-24 09:15:03	202.92.235.0/24	9498	🟩	🟩	🟩	🟩
1488	2011-08-24 09:15:33	83.223.224.0/20	31733	🟩	🟩	🟩	🟫
1489	2011-08-24 09:15:33	92.114.90.0/24	41273	🟩	🟩	🟩	🟩
1490	2011-08-24 09:15:33	92.114.88.0/24	41273	🟩	🟩	🟩	🟩
1491	2011-08-24 09:17:35	2.93.163.0/24	8402	🟩	🟩	🟩	🟩
1492	2011-08-24 09:18:05	124.16.248.0/22	7497	🟩	🟩	🟩	🟩
1493	2011-08-24 09:18:05	180.86.56.0/24	17964	🟩	🟩	🟩	🟩
1494	2011-08-24 09:19:36	81.181.20.0/24	29171	🟩	🟩	🟩	🟩
1495	2011-08-24 09:19:36	212.80.26.0/23	44889	🟩	🟩	🟩	🟩
1496	2011-08-24 09:22:07	213.16.48.0/24	8866	🟩	🟩	🟩	🟩
1497	2011-08-24 09:22:38	190.220.136.0/21	11664	🟩	🟩	🟩	🟩
1498	2011-08-24 09:22:38	200.59.160.0/24	22699	🟩	🟩	🟩	🟩
1499	2011-08-24 09:22:38	200.80.205.0/24	27711	🟩	🟩	🟩	🟩
1500	2011-08-24 09:23:40	203.26.77.0/24	18212	🟩	🟩	🟩	🟩
1501	2011-08-24 09:25:42	196.32.224.0/24	37178	🟩	🟩	🟩	🟩
1502	2011-08-24 09:27:44	190.103.250.0/24	27983	🟩	🟩	🟩	🟩
1503	2011-08-24 09:27:44	190.103.238.0/24	27983	🟩	🟩	🟩	🟩
1504	2011-08-24 09:27:44	190.185.145.0/24	27983	🟩	🟩	🟩	🟩
1505	2011-08-24 09:27:44	190.103.249.0/24	27983	🟩	🟩	🟩	🟩
1506	2011-08-24 09:27:44	190.103.239.0/24	27983	🟩	🟩	🟩	🟩
1507	2011-08-24 09:29:17	66.170.44.0/22	16698	🟩	🟩	🟩	🟩
1508	2011-08-24 09:29:17	66.249.232.0/22	16698	🟩	🟩	🟩	🟩
1509	2011-08-24 09:29:17	66.249.228.0/22	16698	🟩	🟩	🟩	🟩
1510	2011-08-24 09:30:17	212.16.89.0/24	44889	🟩	🟩	🟩	🟩
1511	2011-08-24 09:30:17	97.107.48.0/20	29953	🟩	🟩	🟩	🟩
1512	2011-08-24 09:31:49	91.212.106.0/24	12695	🟩	🟩	🟩	🟩
1513	2011-08-24 09:31:49	92.38.210.0/23	12695	🟩	🟩	🟩	🟩
1514	2011-08-24 09:31:49	95.163.64.0/19	12695	🟩	🟩	🟩	🟩
1515	2011-08-24 09:31:49	95.163.96.0/19	12695	🟩	🟩	🟩	🟩
1516	2011-08-24 09:31:49	95.163.0.0/19	12695	🟩	🟩	🟩	🟩
1517	2011-08-24 09:31:49	178.170.234.0/23	12695	🟩	🟩	🟩	🟩
1518	2011-08-24 09:31:49	89.208.136.0/21	12695	🟩	🟩	🟩	🟩
1519	2011-08-24 09:31:49	89.208.144.0/20	12695	🟩	🟩	🟩	🟩
1520	2011-08-24 09:31:49	89.208.32.0/19	12695	🟩	🟩	🟩	🟩
1521	2011-08-24 09:31:49	79.137.208.0/20	12695	🟩	🟩	🟩	🟩
1522	2011-08-24 09:31:49	79.137.224.0/20	12695	🟩	🟩	🟩	🟩
1523	2011-08-24 09:31:49	79.137.160.0/19	12695	🟩	🟩	🟩	🟩
1524	2011-08-24 09:31:49	79.137.128.0/23	12695	🟩	🟩	🟩	🟩

Color Scale for Feature Scores

-1.093 26.731

2011-08-24 08:00:02 Prefix: 83.223.224.0/19

15469-CH 15576-CH 6772-CH (3356-US) (1299-EU) 12182-US 31733-RU

2011-08-24 09:15:33 Prefix: 83.223.224.0/20

15469-CH (9002-GB) (9002-GB) (9002-GB) (9002-GB) (9002-GB) (6854-RU) (6854-RU) (6854-RU) (6854-RU) 42632-RU 48347-UA 43659-RU 31733-RU

Figure 2.14: **BGP Event Visualization**: Using geo location to visualization BGP routes.

2.2.3 BGP Features Visualizer

BGP Features Visualizer was introduced in Deliverable 4.1 for the purpose of visualizing and analyzing the features presented in Sections 2.1.3, 2.1.4 and 2.1.5, and allowing the analyst to perform BGP hijack detection and attribution. BGP Features Visualizer is comprised of three visualization components: 1) Parallel Coordinates User Interface, 2) Feature Graph view, and 3) Combined Graph view. Each one of these components will be described in detail in the paragraphs that follow.

The *Parallel Coordinates User Interface* uses the Parallel Coordinates visualization approach to visualize the interrelationships between the different features, which is further enhanced with filtering capabilities. Using this view and combined with the graph views, the analyst can perform visual fusion of the aforementioned features for the purpose of BGP path or MOAS hijacking detection.

Each Parallel Coordinate represents an individual feature, while the red lines that run across the Parallel Coordinates represent the values of each individual BGP event. It should be noted that the direction of each parallel coordinate is such that the extreme values of interest, are positioned at the upper part of the Parallel Coordinates User Interface, i.e, low values of the CAP and $CAPZ$ features and high values of the CGL, $CGLZ$, *GDO* and *eccentricity* features are positioned at the upper part of the view. Using this view, the analyst can immediately gain an understanding of the distribution that each feature follows, as well as possible correlations that may exist between them. Furthermore, the outlier values of each feature, which are of particular importance in the analysis, are easy to detect.

As it was mentioned earlier, the analyst is interested in events that have extreme values in all the available features at the same time. The disadvantage of the Parallel Coordinates visualization, is that due to high cluttering, it does not facilitate the means necessary to accomplish this goal. In order to enhance the Parallel Coordinates approach, filtering is employed. More specifically, sliders are attached onto each parallel coordinate, whose position represents the value of the corresponding threshold. The events whose at least one feature value is below the predefined thresholds are omitted from the visualization. Thus, by adjusting the position of the thresholds the analyst can perform visual fusion of the available features, so as to focus only on the most interesting events. The definition of the thresholds has a direct impact onto the feature and combined graph views, that directly reflect the remaining events, as well as any relationships between them.

Feature Graph view provides a graph based visualization of each feature. Using this view the analyst can detect which ASes and Countries are involved in suspicious events according to the selected feature, as well as any relationships that may exist between them. The nodes of the graph represent either ASes or Countries, while the existence of an edge between an AS and a Country node implies the occurrence of a BGP event (MOAS or path alteration).

As it was mentioned earlier, the analyst is interested only in the extreme values of each feature, thus, the width of the edges is proportional to the importance of the value of the corresponding feature. For example, in the possibility feature graph (CAP), high edge widths represent low feature values, while in the geographic deviation feature graph (CGL), high edge widths represent high feature values.

Each feature graph is comprised of two types of views. The classical graph view which represents a graph directed layout and the world map view which positions the country nodes in their relative position on the globe, while the ASes are positioned close to their country of origin. Thus, the analysts can perform his/her analysis by also incorporating the geographical aspects of the BGP events.

Finally, the *Combined Graph view* is a fused graph of all the individual feature graphs that is used for the purpose of highlighting their structural similarities. These similarities are used to highlight the suspicious BGP activities across any number of features, as well as reveal possible participation of an actor in multiple events, visible from multiple features. One important aspect of the *Combined Graph view* is that it emphasizes the events that exist across many features by increasing the width of the corresponding edge, as well as changing its color. Thus, the visual features of each edge capture the number of features the corresponding event is visible from, after the application of filtering.

Thus, by utilizing the three views provided by BGP Features Visualizer, the analyst can perform visual fusion of the aforementioned features, so as to navigate through the bulk of BGP activity and detect BGP hijacking events in an intuitive manner.

2.2.4 BGP Routing Changes Graph

BGP Routing Changes Graph was presented in Deliverable 4.1 and aims at providing a hierarchical graph visualization scheme to present BGP routing changes caused by the BGP announcements and withdrawals. The purpose of this visualization is to allow the analyst to have an overview of the BGP activity caused by the BGP messages and detect any substantial changes that need further investigation and root cause identification.

The initial AS-graph utilized by BGP Routing Changes Graph represents the Internet topology at an Autonomous System (AS) level. In other words, the nodes of the AS-graph represent ASes and the edges physical connections between them. The visual features of the nodes and the edges represent the amount of routing change observed at a specific user defined time window. More specifically, the width of the edges represents the amount of routing change serviced by the corresponding edge, while the size of the nodes represents the amount of prefix ownership change of the corresponding AS. Routing change and prefix ownership change are both measured in IP addresses. Furthermore, color is utilized to show the sign of the change, red for negative and green for positive changes.

A hierarchical clustering scheme is applied on the AS-graph in order to create a series of course to fine graphs and facilitate the graph visualizations by providing a scalable hierarchical method. The hierarchical clustering scheme creates multiple layers or levels, each one having different granularity. Each node of a coarse graph is a cluster comprised of multiple ASes, while each edge of the graph is a fusion of edges from the AS-graph. Thus, the analyst is presented with an overview of the BGP activity shown in a relatively small graph, while she/he is also able to focus on specific parts of the graph that may be of interest, by expanding the clusters into the ASes of which they are comprised.

A novel contribution of the BGP Routing Changes Graph visualization is the definition of a method capable of optimizing the mapping procedure from the input dataset to the visualization space. Thus, more information is presented to the user, enabling him/her to take more informed decisions and capture more patterns. To achieve this, the information content of the visualization is measured using entropy measures. The exact amount of information currently displayed is subject to the screen parameters(e.g. size, or resolution) and the mapping function used (e.g. linear, or logarithmic). For a specific display, this value depends only on the mapping function utilized. This suggests that it is possible to analytically define the mapping function so as to maximize the information displayed by the visualization. In the case of BGP Routing Changes Graph, the downhill simplex method [53] is applied in order to find the optimum mapping function, which maximizes the information content of the graph visualizations.

3 Validating Hijacks

This Chapter discusses the feasibility and challenges faced in the validation of prefix hijackings. The first issue to clarify is what is understood by “prefix hijacking” and “validation”. By looking at the literature, we first establish what has been done so far in order to validate the results, and then explain why these steps are not sufficient. We look at the lack of owner feedback and explain the reasons behind them, and at the lack of information in BGP messages. We then present methods that help circumvent these issues, and help us form an informed position on a prefix’s situation.

3.1 Common Challenges in Identifying Prefix Hijacking

3.1.1 Introduction

Over the last couple of years, much work has been carried out in order to detect prefix hijackings. Traditionally, prefix hijacking happens when an ASN originates a prefix that it does not own. This definition is very broad and encompasses situations that can result from router misconfiguration (detailed in [28]), downright incompetence (e.g. the YouTube/Pakistan Telecom hijack [37]), or attacks against the routing infrastructure. The two first categories can cause result in large-scale outages and are typically short-lived. In the context of VIS-SENSE – and particularly in the context of WP6 – we focus on hijacks that appear to have been used as a stepping launch to other malicious ends, like spam. These are *malicious* hijacks. Their security impact is high because they potentially enable an attacker to steal IP identities. In the case of spam, this allows the attacker to circumvent traditional IP blacklist based on IP reputations. In any case, identifying the responsible party becomes even harder, and legal actions could mistakenly be taken against the owner of the IP block.

Validating hijack cases following router misconfigurations, is not too complicated since it heavily impacts the service of the owner. Post-mortem analysis, such as [37] can easily be carried out. However, Khare et al. [24] systematically analysed BGP data for (non-malicious) hijacks between 2003 and 2010, and most of the events they uncovered were not publicly known or analyzed; although the authors suggest the prefix owners already knew about them.

3.1.2 Availability of Ground-Truth Information

Previous work always relies on owner feedback in order to validate, or invalidate a hijack. PHAS [27] did not explicitly seek validation because its operative design goal was to report origin changes to interested prefix owners. These owners were then in charge of making the decision on whether this routing state is legitimate or not, and to take the appropriate actions. More recently, Shi et al. [42] made considerable efforts in order to validate the output of Argus by analysing the network circumstances that led to prefixes being alerted, and also by contacting the prefix owner.

The reason owner feedback is systematically sought is that differentiating between a false positive and a hijack attack is a non-trivial task because of a lack of availability of ground-truth information. The owner of the prefix is the only one who knows – without doubt – what the expected behaviour of the prefix is. This information includes which prefix flavours are supposed to be announced and by whom. Third-party observers generally have no access to this information. WHOIS databases, including RPSL route objects are notoriously out of date, and unmaintained [33, 42]. This is partly explained by the fact that peering information is oftentimes considered as sensitive operational data by corporations, and thus remain undisclosed [40]. Consequently, the content of registration databases cannot be used to confirm or infirm a suspect situation. The question to answer before interpreting WHOIS data is whether it is customary for the owner of the prefix of interest to keep the records up to date. This can be done with the help of the last-modified timestamps associated with any of the WHOIS database entries.

Another consequence of this lack of ground-truth is the impossibility of using machine learning techniques in order to detect prefix hijacking. This is because it is not possible to train the algorithm over a truth set, containing the announcements for a prefix as well as their validity.

3.1.3 Imperfection of BGP Data

On top of the problems for accessing ground-truth information described in the previous Section, Roughan et al. [40] detail several weaknesses in BGP data. This section summarizes the comments that would be relevant to BGP hijacking.

BGP is an information-hiding protocol per design: networks are able to exchange routing information without revealing anything about their own internal structure. As a result, information such as the size of the network (e.g. the number of border points with neighbouring ASes) is unknown. As discussed previously, the peering policy with these neighbours also remains hidden. This leads to the abstraction that an AS is an atomic node, which is an over-simplification as it lacks policy diversity and multi-connectivity between multiple ASes. This multi-connectivity is important because large ASes – spanning countries, sometimes even continents – may not have the same view of the Internet depending on the location of a machine within that network. In general, the belief that the Internet can be efficiently and correctly modeled into a digraph leads to a global over abstraction of the network.

Moreover, in order to remain scalable over the whole Internet, only the best selected path is propagated. As a result, the data lacks route diversity. Since BGP is a policy-based routing protocol, this best path is not necessarily the best one in terms of topology. Moreover, many paths appear not to be forwarded far from network edges. Back up links, for example, seem to appear in the wild only when there is a major issue at the network edge. As a result normality models often used by prefix hijacking detection techniques are unable to react to this kind of event in an appropriate manner.

Route collectors peer with a large number of distinct peers, and are geographically diverse. However, as previously mentioned, a large AS may not have the same view of the Internet depending on its peering point. As a result, the reported forwarded routes by this kind of peer is really dependant on the collector location. Route collectors are geographically diverse in order to ensure better reach within the network. However, they show a heavy bias toward core networks because they are often located within IPXs and/or peering with backbone networks.

Finally, routing security researchers often simplify the model of BGP business relationships by classifying peering agreements into either provider-customer, peer-to-peer, and siblings. This classification is often coupled with the expectation of valley-free paths [19]. However, valleys appear to be more of a rule than an exception in BGP routes because the peering agreements cannot be as neatly classified as the previous three relationships. As a result, when checking against a peering policy-based model, either a wrong policy is inferred by the algorithm, or an alert is raised due to a policy violation. In both cases, it is the result of an over-simplification of the reality which may lead to serious error of judgements when reviewing candidate hijack events.

3.2 Possible Validation Techniques

In order to overcome the limitations due to a lack of ground-truth information, we propose several techniques through which an informed opinion can be made on the maliciousness of an event.

3.2.1 Validation through Topology

The long-term evolution of the topology graph of a given suspect prefix – both before and after is has been suspected – can help give insight on the root cause of an event during a post-mortem analysis.

By observing the way the prefix was being announced before an alert was raised, as well as after the alert has been discarded, a long-term BGP view of the prefix can be created. This view informs us of the long-term topology implication of the event.

For example, if a network is changing its upstream provider, it is expected to present a MOAS and a path anomaly at a certain point in time. However, observing the BGP behaviour on the long-term would underline that the network has effectively stopped peering with its upstream in order to go to a new one. This, then, retroactively invalidates the hypothesis of a hijack.

3.2.2 Validation through Algorithmics

The VIS-SENSE consortium introduced and implemented a set of distinct algorithms that look at BGP data in a very different way, which were presented in the deliverables related to WP3 and reviewed in Chapter 2. These algorithms all focus on a single property (or a specific set of properties) of a prefix announcement that could be interpreted as the side-effect of an attack on the routing infrastructure. However, these situations might also be the result of a legitimate announcement, thus raising a false positive. By combining the multiple features behind the different algorithms, we build a multi-sided view of the a single prefix. Combining these views help us infer the effect of each feature, and strengthens (or weakens) the suspicion behind a prefix.

For example, a prefix exhibiting spam traffic and a geospatial incoherence is more suspicious than a prefix that exhibits only any of those features. This is because this combination of features represents a signature more specifically tailored at networks that are hijacked in order to send spam messages. However, there is no guarantee that the prefix is indeed misbehaving – in terms of BGP hijacking – but leads to a series of arguments upon which to judge if the prefix has indeed been abused via BGP means.

3.2.3 Validation with Security Feeds

The conjecture used in the VIS-SENSE project, mainly put forward by Ramachandran et al. [34], is that attackers abuse the global routing infrastructure in order to stealthily carry out illicit activities on the wide Internet. As a results, it is important to use security feeds as inputs in order to know if a prefix has been used in order to carry out malicious activities. VIS-SENSE relies on SpamCloud (Section 1.2.2) in order to detect spam messages.

If there is no record in SpamCloud of any spam being sent from the suspected network, there are two possible outcomes. Either nothing is going on with the prefix, and the detection techniques raised a false positive, following an event due to a legitimate route alteration. As a result, the techniques should be improved in order to avoid raising this kind of alert, if it is repeated. Either the prefix has been hijacked, but no spam was emitted from the stolen space; or no spam emitted from the hijacked space hit a spamtrap maintained by Symantec.cloud. On the other hand, the situation in which SpamCloud received spam from the suspect prefix is not synonymous of a hijack since many networks emit spam daily without having been hijacked.

3.2.4 Validation through the Owner

Contacting the owner of the prefix upon detection of a suspicious event is usually helpful into gaining insight on that particular event. This is usually the standard method for evaluating detection techniques. It has been used extensively by [51, 11, 42, 24].

Unfortunately, it is also not an effective solution because owner's reply ratio can be quite low [11, 42]. This is explained by the fact that some network consider routing information to be private, and, as a result, are not willing to communicate on that topic. Moreover, NOCs are usually quite busy, and have little time to spare. If the problem is finished, they most likely moved on to another matter. Finally, language can also be a boundary: NOCs do not necessarily on the English language in day-to-day operations. As a result, an incoming e-mail written in English, which details operational information could be considered as spam or scam. Moreover, the recipient of the e-mail may not be fluent enough in order to grasp the contents of the email, let alone to be able to reply to it. Sometimes, it may also be that the owner is not aware that their prefix has been abused [27]. In this situation, it is impossible for them to confirm anything.

On top of these communication problems, there is also an ethical issue. Identifying the owner of a prefix is not as straightforward as it seems, because information contained in WHOIS records is not guaranteed to be accurate. Moreover, it is possible that these entries have been altered, or that the contact email address and/or phone number are now under control of attackers, and not of the legitimate owners anymore. This would result in not harmlessly contacting the owners of the prefix, but someone posing as the owner, thus informing them that they have been caught. As a result, it is possible that the attackers would alter their modus operandi in order to avoid being detected.

3.2.5 Validation through Complaint

Sometimes, a network owner notices that something is wrong with their network and wants to reach out to the network community in order to ask other network help in order to block the announcement and/or reach the misbehaving entity. One of the most effective ways to reach out to the community is to use the NANOG (North-American Network Operators' Group) mailing list [3]. By actively NANOG, and other popular mailing lists, some ground-truth information can be gained due to its public disclosure by the owner. Moreover, the content of these mailing lists are archived and available on HTML webpages. A simple websearch can thus be used in order to isolate possible discussions about a particular prefix's announcement.

3.2.6 Discussion

No technique mentioned in this Section, apart from those actively involving the owner, can help effectively reach a conclusion on whether a prefix has been hijacked or not. The idea is to use these techniques as a set of elements to check in order to gain insight on a single case. By gathering as many elements as possible from many different perspectives, a global consensus on an event can be obtained, and then an informed decision on the malicious factor of the event can be decided.

In this context, visualizations play a big role in grasping the different sides of a single story: a large quantity of data is aggregated in a specific way in order to highlight a specific view of the event. By using multiple visualizations, these different sides can timely be checked, without the need to spend resources in analyzing a large amount of logs, binary BGP messages, . . .

3.3 Use Case

This Section presents the abstract use case of the VIS-SENSE framework that enabled the detection of the cases presented in Chapter 4. The following presentation is reminiscent of what was done in D1.2 to present the use cases that lead the development of the framework.

Alice is knowledgeable and fears that BGP can be abused by malicious parties. She is particularly interested in studying the correlation between spam and routing events. To this effect, the VIS-SENSE framework provides her with a unique set of datasets and analysis tools.

Alice's experience taught her that BGP monitoring algorithms are usually create too many alerts in order to manually go through each of them. As a result, she first creates a complex signature that she thinks will match only very peculiar routing events, i.e. events that have a very low chance of resulting from commonly standard routing practices. This signature combines the different VIS-SENSE framework data sources in order to correlate their data. In order to apply the signature to the data sources, she quickly creates a script that makes use of the available WAPI servers that are coupled with every data sources.

The output of her script gives her a list of the prefixes and timestamps associated to the events she imagined were suspicious and that she now wishes to review. In order to investigate those, she makes use of other VIS-SENSE tools that are at her disposal. She decides to open one visualization tool in order to skim through most of the list. During her skimming, an entry peaks her attention. In order to learn more about this event of interest, she decides to check another perspective of the framework for the same event.

Eventually, she checked everything that she thinks is relevant for this event. At this point, she either decides that the event is benign, and resumes her skimming via the original visualization tool, or she decides that she wants to learn more about the involved networks. In order to do so, she checks the routing history with the framework, checks the WHOIS entries through the help of RIR databases, and uses a web-search engine to look for other report of malicious activity about that prefix, for example in the NANOG mailing list. Once these sources of information have been checked, she decides if the event is benign or not. If she thinks it is a real case of hijack, she considers getting in touch with the owner of the prefix in order to gain further information about the situation.

4 Experimental Results

This Chapter presents three distinct cases where spammers are believed to have used BGP in order to stealthily emit spam into the network. The first Section focuses on the Link-Telecom hijack case which took place between April and August 2011 where an American spammer hijacked the prefix of a Russian regional ISP. The second Section focuses on cases of fly-by spamming that occurred during the first half of 2013 where spammers abused IPv4 black space in order to send spam. Finally, the third Section focuses on the Bulgarian case, where a spammer was believed to have hijacked a set of a Bulgarian ISP's subprefixes in order to abuse them with spam and scam activities.

These studies make extensive use of the VIS-SENSE framework as data sources and detection algorithms. Moreover, special care is taken in order to take all the steps in order to reach a definitive validation on the cases. If this verdict could not be reached, arguments for and against the hijack scenario are presented and show the limits of what state-of-the-art techniques can achieve today in terms of validation prefix hijacking.

4.1 Case Study I: The Link-Telecom Hijack

One of the VIS-SENSE use cases presented in Deliverable 1.2 "Use Case Analysis and User Scenarios" is related to the identification of BGP hijacks performed by spammers to send spam from the stolen IP space, hinder traceability and remain stealthy. In this section we present the visual analysis of a validated case of malicious BGP hijack performed by a spammer in order to use the stolen IP space to send spam. While anecdotal evidence of spam correlated with short-lived BGP routes (i.e., lasting less than a day) to unadvertised networks have already been reported in [34, 22], the case described here after regards the hijack of unadvertised prefixes for no less than 5 months during which the stolen IP space was used for sending spam. This hijack case is further described and analysed in [9, 48, 12, 17] as well as in Deliverables 3.2 "Correlation analysis and abnormal event detection module" and 4.1 "Visual Network Analysis".

First we present the identification of the hijack event in the data-plane traces collected by SpamTracer using the VisTracer visual analytics tool. By correlating this event with the other BGP hijack detection tools through the VIS-SENSE framework, we can further analyse the event using the BGP Features Visualizer, this time from a purely control-plane point of view. The malicious nature of this event suggested by our combined analysis was later confirmed by the owner of the victim network when he complained on a public mailing list that his network had been hijacked.

4.1.1 Visual Analysis with VisTracer

From the *ASN Overview* (Figure 4.1), 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).

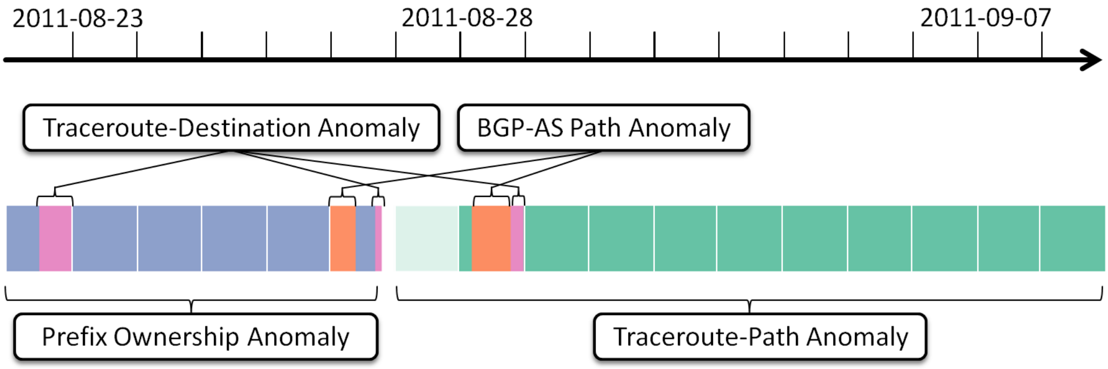


Figure 4.1: The *ASN Overview* of AS31733 reveals many different anomalies over a longer period of time.

Figure 4.2 presents the *Target History Visualization* of a monitored host within AS31733 exhibiting a combination of Traceroute Anomalies and BGP AS Path Anomalies on August 29th. The *Target History Visualization* 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 indicating a major routing change related to AS31733 occurred. By looking at the anomalies extracted for that case, we can also see that all anomalies were observed on a particular day, i.e., at the same time as 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. We can further correlate this routing change with the fact that traceroute paths used to traverse ASes located in the US before the change instead of Russia after the change. Lastly the routing change can also be linked with the monitored host and AS to become unreachable.

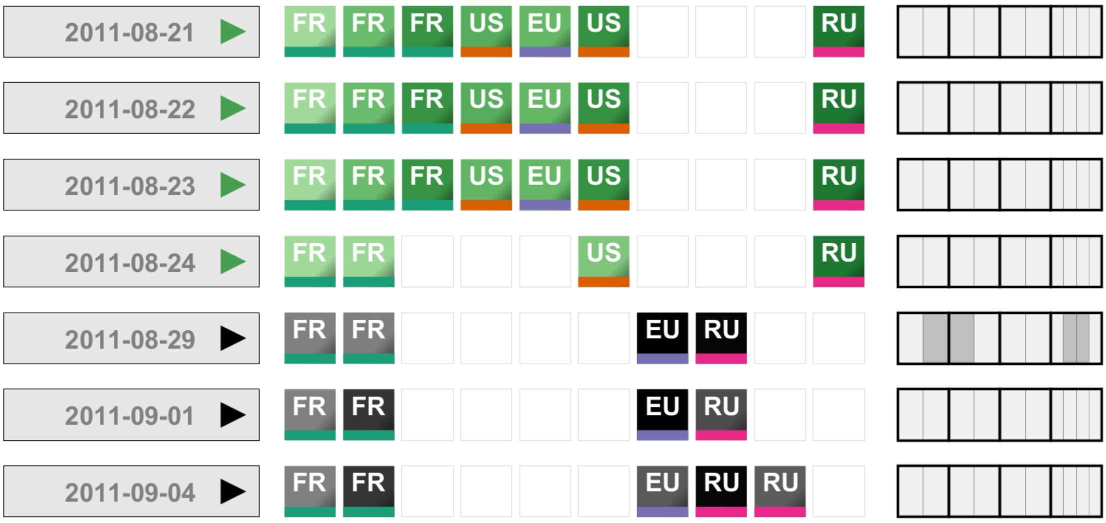


Figure 4.2: The *Target History Visualization* shows the significant difference in the *set* of ASes traversed between the fourth and fifth day. The routing anomalies observed are also shown.

Figure 4.3 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.

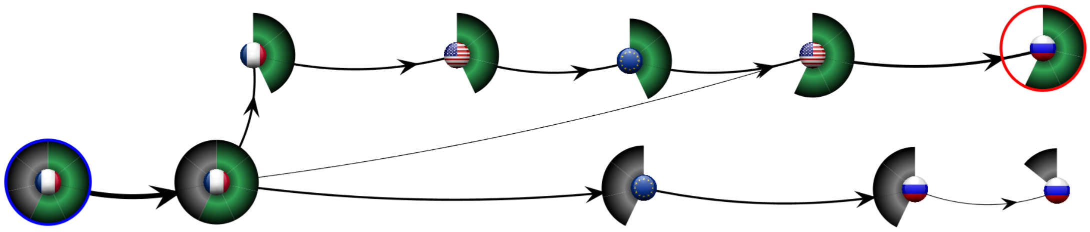


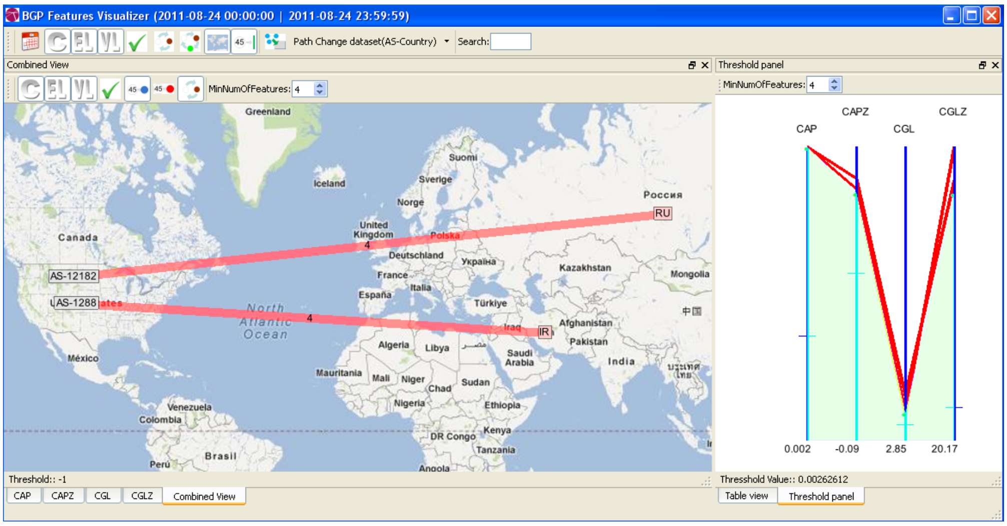
Figure 4.3: The *Graph Visualization* shows the significant difference in the *sequence* of ASes traversed. It also highlights the unreachability of the destination AS after the routing change occurred.

On both August 25^{th} and August 29^{th} 2011 changes were observed in the traceroutes and BGP routes toward 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.

4.1.2 BGP Features Visualizer

Figure 4.4 depicts the BGP Features Visualizer approach (Deliverable 4.1 and Section 2.2.3) visualizing all the path change events that occurred on 24-Aug-2013. Filtering has already been applied in order to allow the analyst to focus only on the most suspicious events from the bulk of BGP path change phenomena. Furthermore, figure 4.4 shows the combined graph view and the Parallel Coordinates User Interface. There exist two types of nodes, the Country nodes, which host the prefixes under investigation, and the AS nodes, which represent suspicious ASes. An edge between these types of nodes corresponds to a path change event, for the path toward the corresponding Country, while the AS node depicts the most suspicious AS for this path.

Each path change event is characterized by four features: probability of intermediate country appearance (CAP), z-score of the probability of intermediate country appearance



BGP Features Visualizer (2011-08-24 00:00:00 | 2011-08-24 23:59:59)

Path Change dataset(AS-Country)

Search:

Combined View

MinNumOfFeatures: 4

Threshold panel

MinNumOfFeatures: 4

Threshold:: -1

Thresshold Value:: 0.00262612

Figure 4.4: The BGP Features Visualizer visualization of all the path change events that occurred on 24-Aug-2013, after the application of filtering. The Link-Telecom case AS12182 connected to Russia is one of the top two suspicious path change events. This figure depicts the combined graph view and Parallel Coordinates User Interface.

(CAPZ), geographic deviation induced by an intermediate country (CGL), and the z-score of the geographic deviation induced by an intermediate country (CGLZ). Low CAP and CAPZ, and high CGL and CGLZ values indicate possible BGP path change anomalies that need further investigation. Thus, using the Parallel Coordinates User Interface, the analyst is able to set specific feature thresholds and focus on the most suspicious cases.

The two most interesting BGP path change events that took place on 24-Aug-2013 are depicted in figure 4.4. One of these events, corresponds to the known Link-Telecom hijack event (link from AS12182 located in USA to Russia). The BGP Features Visualizer approach detected this event in an intuitive manner using the expert knowledge to set the filtering thresholds and interpret the results.

4.1.3 The Complete Story

On August 20th 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 [7]. 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. As a countermeasure, Link-Telecom responded on August 24, by announcing more specific prefixes of the hijacked prefixes in order to regain control over the hijacked IP space. 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.

This case study shows the complementarity of the developed BGP hijack detection approaches and visualisations in the identification and investigation of suspicious hijack events. It further shows the added value of correlating suspicious events from different tools to focus on the investigation of the most suspicious and likely malicious ones.

4.2 Case Study II: Fly-By Spammers

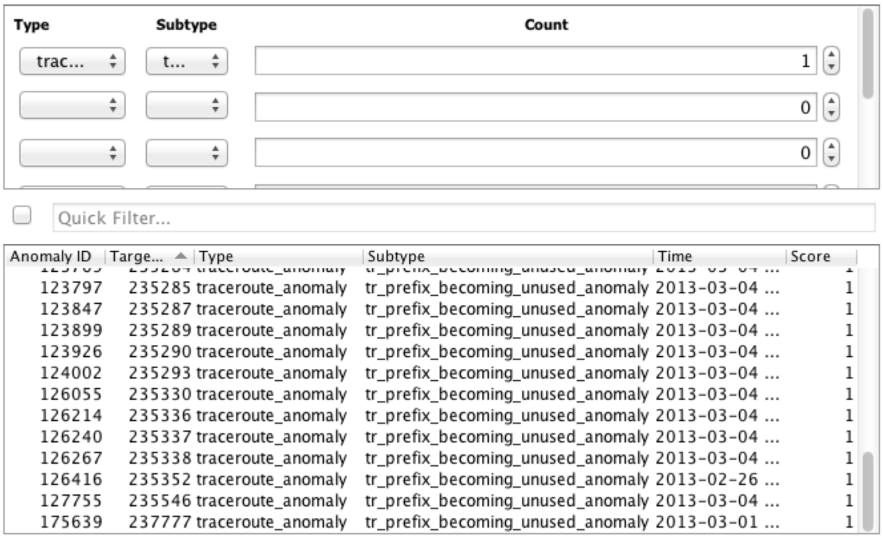
The first case study presented here here above was related to a long-lived hijack performed by a spammer in order to launch spam campaigns from the stolen IP space. In this section we present another visual analysis of hijacked unadvertised network blocks correlated with spam received at Symantec.cloud spamtraps. Unlike the hijack of the first case study, hijack cases presented here were rather short-lived, i.e., they lasted at most 20 days and mostly less than 5 days. Those spammers are thus able to take control of IP address space in order to send spam from clean, non-blacklisted IP addresses. We commonly refer to spammers performing short-lived hijacks as *fly-by spammers*.

Similar to the first case study, we first present the identification of the hijack events using SpamTracer and the VisTracer visual analytics tool. We then further analyse those events using the BGP Features Visualizer.

4.2.1 Visual Analysis with VisTracer

Fly-by spammer cases are originally characterised by spam networks being advertised in the Internet for a short period of time. In order to focus our visual analysis on these cases we take advantage of the SpamTracer routing anomalies filter panel (see figure 2.13 (2) in section 2.2.1 page 40) which allows the analyst to consider only cases matching a given combination of routing anomalies (type, e.g., Traceroute Anomaly and subtype, e.g., Traceroute Prefix Becoming Unused Anomaly), corresponding for instance to a specific routing behavior he is willing to analyse. In the case of fly-by spammer hijacks we are interested in networks monitored by SpamTracer exhibiting at least a

Traceroute Prefix Becoming Unused Anomaly indicating that a network stopped being advertised during its monitoring period, i.e., it disappeared from the routing tables of queried BGP collectors and traceroutes all failed to reach the destination network and host.



Type	Subtype	Count
trac...	t...	1
		0
		0

□ Quick Filter...

Anomaly ID	Targe...	Type	Subtype	Time	Score
123797	235285	traceroute_anomaly	tr_prefix_becoming_unused_anomaly	2013-03-04 ...	1
123847	235287	traceroute_anomaly	tr_prefix_becoming_unused_anomaly	2013-03-04 ...	1
123899	235289	traceroute_anomaly	tr_prefix_becoming_unused_anomaly	2013-03-04 ...	1
123926	235290	traceroute_anomaly	tr_prefix_becoming_unused_anomaly	2013-03-04 ...	1
124002	235293	traceroute_anomaly	tr_prefix_becoming_unused_anomaly	2013-03-04 ...	1
126055	235330	traceroute_anomaly	tr_prefix_becoming_unused_anomaly	2013-03-04 ...	1
126214	235336	traceroute_anomaly	tr_prefix_becoming_unused_anomaly	2013-03-04 ...	1
126240	235337	traceroute_anomaly	tr_prefix_becoming_unused_anomaly	2013-03-04 ...	1
126267	235338	traceroute_anomaly	tr_prefix_becoming_unused_anomaly	2013-03-04 ...	1
126416	235352	traceroute_anomaly	tr_prefix_becoming_unused_anomaly	2013-02-26 ...	1
127755	235546	traceroute_anomaly	tr_prefix_becoming_unused_anomaly	2013-03-04 ...	1
175639	237777	traceroute_anomaly	tr_prefix_becoming_unused_anomaly	2013-03-01 ...	1

Figure 4.5: The *Routing Anomalies* filter panel allows to consider networks exhibiting a given combination of routing anomalies. In this case we want to focus on fly-by spammer cases so we select networks exhibiting a *Traceroute Prefix Becoming Unused Anomaly* indicating a spam network likely advertised by a fly-by spammer for a short period of time.

VisTracer then allows to visualise the *ASN Overview* specifically for the monitored networks found exhibiting a *Traceroute Prefix Becoming Unused Anomaly.* Figure 4.6) shows the *ASN Overview* of an AS involved in a suspicious fly-by spammer hijack. One network in this AS was monitored from February 1^{st} to February 9^{th}. On February 4^{th}, four Traceroute Anomalies were observed: (i) a Hop Count Anomaly (i.e., a major change in the length of traceroutes), (ii) an IP Reachability Anomaly (i.e., the destination host became unreachable), (iii) an AS Reachability Anomaly (i.e., the destination AS became unreachable), and (iv) a Prefix Becoming Unused Anomaly (i.e., the destination network address block disappeared from the routing tables of queried BGP collectors).

Looking closer at the traceroutes collected for the suspicious hijacked AS we can see in the *Target History Visualization* in figure 4.7 that the AS-level traceroute collected on

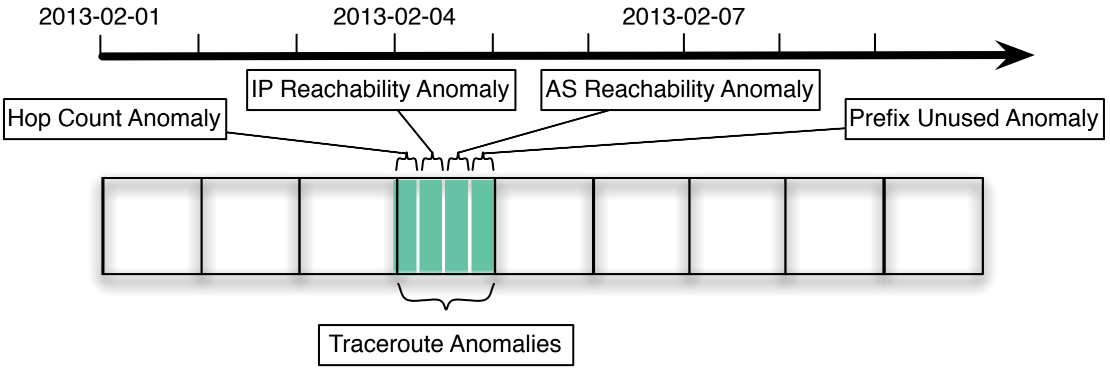


Figure 4.6: The *ASN Overview* of an AS involved in a suspicious fly-by spammer hijack reveals many routing anomalies on February 4^{th} when the hijacked network was released by the spammer and became unadvertised.

the first day managed to reach the destination AS. However AS-level traceroutes collected afterwards suggest the destination network became unadvertised as traceroute paths are very small (i.e., they cross two ASes while the first crossed seven ASes) and stop after two ASes in the path. All this likely results from routers dropping traceroute probes due to the absence of route to the network in BGP. The *Target History Visualization* also shows the correlation between the routing changes observed and the four routing anomalies extracted from the traces on February 4^{th}.

Figure 4.8 presents the *Graph Visualization* of the same monitored host within the suspicious AS. This visualization shows the *sequence* of IP hops and countries traversed by the traceroutes. In this case, we can see from the clock glyphs for each IP hop that on the first day of the monitoring period the traceroute was 12 hops long and reached the destination host. On subsequent days however traceroutes failed to go beyond the four first IP hops. From the sequence of countries traversed we can also observe this bizarre link between the penultimate hop apparently located in Ukraine and the last hop apparently located in the Virgin Islands, two countries located so far away from each others that a legitimate IP-level link between them is very unlikely. This could in fact be the result of the destination network being hijacked and advertised via an ISP in Ukraine, similar to the hijack and the advertisements of Link-Telecom networks via the US ISP Internap.

Evidence collected from the visual analysis of SpamTracer data related to this specific fly-by spammer case allows to conclude that:

- spam was received from the suspicious network;

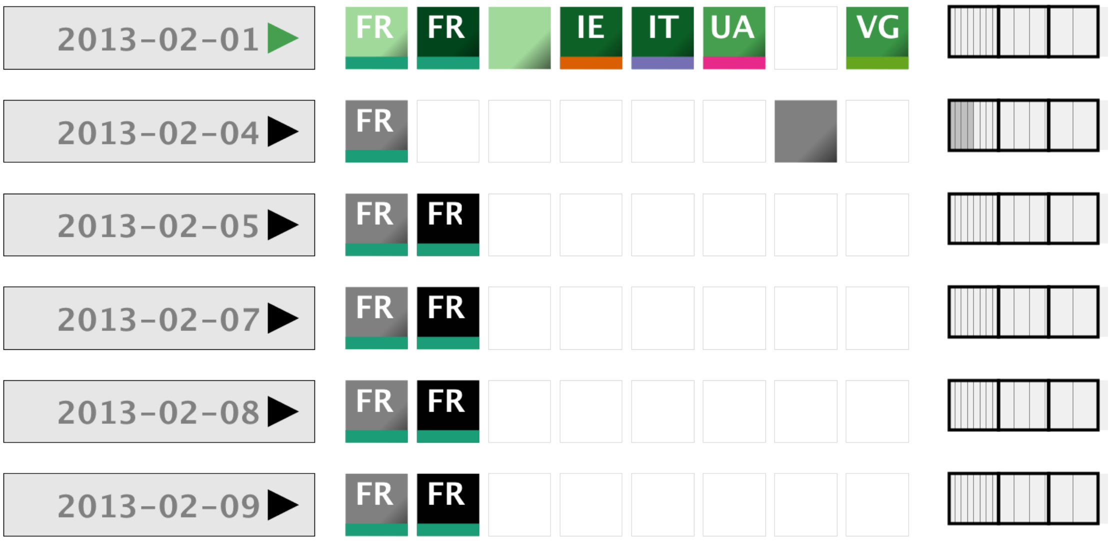


Figure 4.7: The *Target History Visualization* shows the significant difference in the *set* of ASes traversed between February 1st and February 4th. The four routing anomalies observed on the second day are also shown.

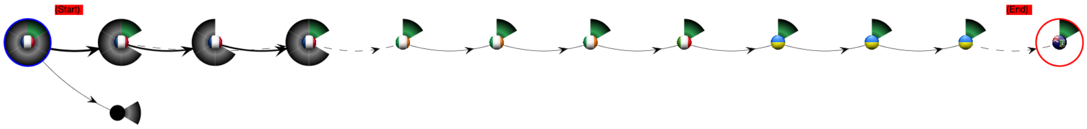
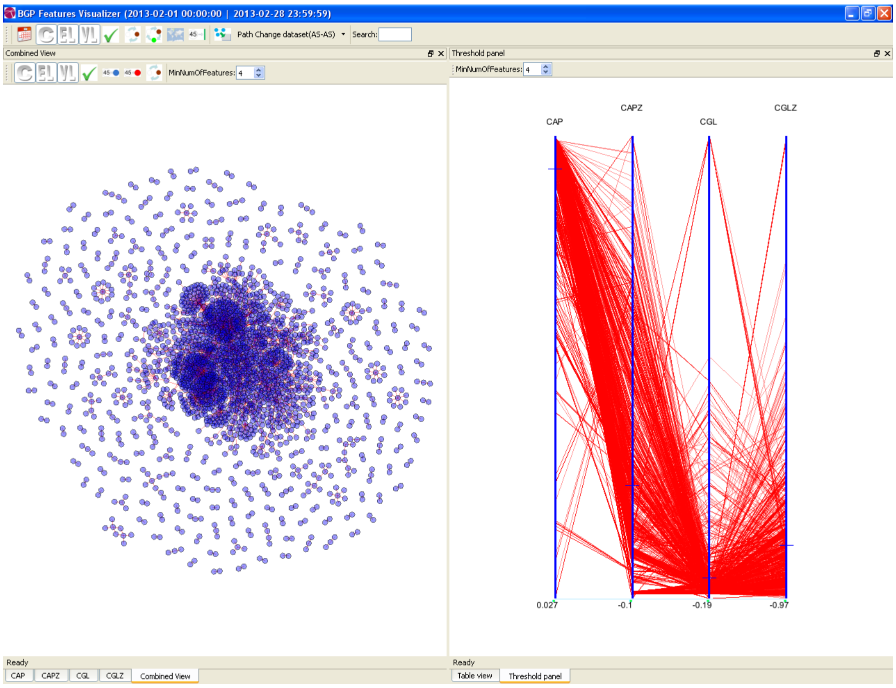


Figure 4.8: The *Graph Visualization* shows the significant difference in the *sequence* of ASes traversed. It also highlights the unreachability of the destination AS after the destination network was withdrawn from the Internet routing tables. Finally, it points out the suspicious IP-level link between the penultimate hop and the last hop apparently connecting a host in Ukraine with one in the Virgin Islands.



MinNumOfFeatures: 4

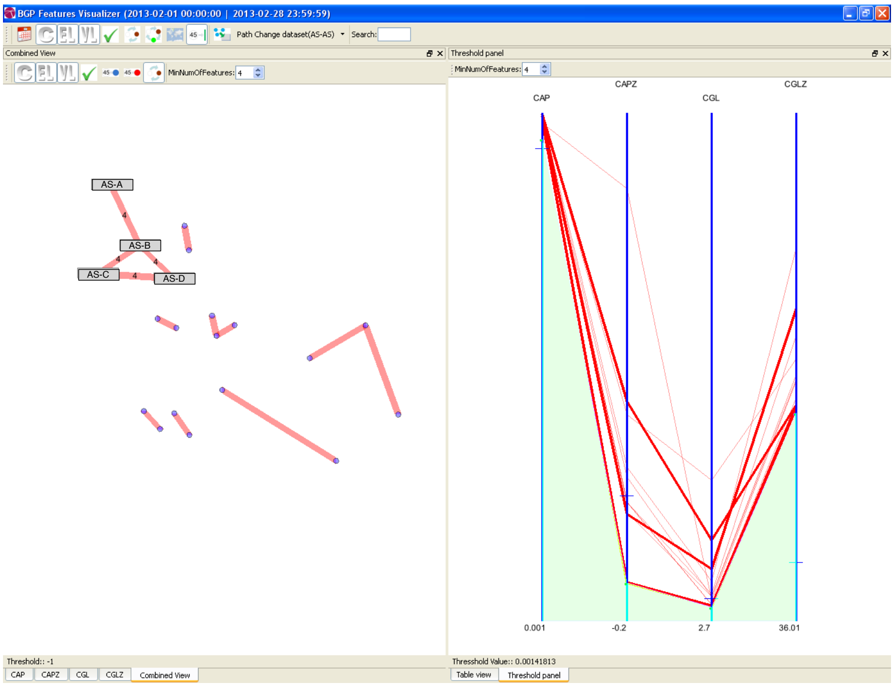
MinNumOfFeatures: 4

Figure 4.9: BGP Features Visualization of all the suspicious BGP announcements that occurred on February 2013. This figure shows the combined graph view and the Parallel Coordinates User Interface.

- the network became unadvertised within a few days after spam was received which is the expected routing behavior of networks hijacked by fly-by spammers, and;

- the network apparently located in the Virgin Islands was suspiciously advertised via an Ukrainian ISP reinforcing the idea of the network being hijacked.

While we particularly focused our visual analysis on a given case we will show in the next two sections that the routing behavior characterised here is also observed in many other networks we suspect were hijacked by fly-by spammers.



BGP Features Visualizer (2013-02-01 00:00:00 | 2013-02-28 23:59:59)

Path Change dataset(AS-AS)

Search:

MinNumOfFeatures: 4

MinNumOfFeatures: 4

Threshold:: -1

Thresshold Value:: 0.00141813

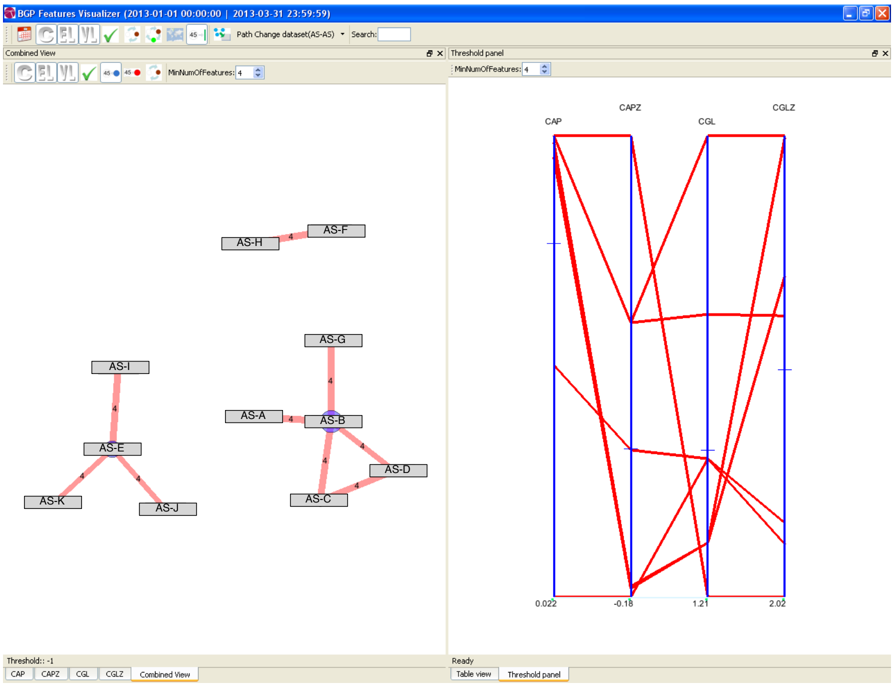
Figure 4.10: BGP Features Visualization of the suspicious BGP announcements that occurred on February 2013 after the application of filtering to further reduce the number of ASes and focus on the most interesting cases. The group of selected ASes, which are also involved in spam activities, is comprised of: AS-A, AS-B, AS-C, and AS-D. This figure shows the combined graph view and Parallel Coordinates User Interface.

4.2.2 BGP Features Visualizer: Investigating the relationship between BGP activity and spammers

4.2.2.1 Visualizing the most Suspicious BGP activity that took place on February

The BGP Features Visualizer approach provides a visual representation of the BGP features defined in Sections 2.1.3, 2.1.4 and 2.1.5 for the purpose of focusing on suspicious BGP activities and preform attack detection and attribution. Furthermore, the graph representations provided by BGP features Visualizer allow for the correlation of the involved ASes and Countries by revealing underling relationships.

The analysis that follows combines two different approaches: 1) the approach of Spam-



BGP Features Visualizer (2013-01-01 00:00:00 | 2013-03-31 23:59:59)

Path Change dataset(AS-AS) Search:

Combined View

MinNumOfFeatures: 4

Threshold panel

MinNumOfFeatures: 4

Threshold:: -1

Ready

Figure 4.11: The BGP Features Visualizer visualization of all the spam prefixes for the period from 01-Jan-2013 to 31-Mar-2013. The depicted ASes are either performing suspicious activity or are owners of the corresponding prefixes. This figure shows the combined graph view and Parallel Coordinates User Interface.

Tracer (Deliverable 3.2) for the analysis of occurrence of spammers that also perform BGP hijacks, and 2) the BGP features presented briefly in Sections 2.1.3, 2.1.4 and 2.1.5 (also in Deliverable 3.2) visualized using BGP Features Visualizer.

To start with, the analysis begins with a possible investigation between the occurrence of spammers with suspicious BGP activity. All the suspicious BGP announcements (Section 2.1.4) that occurred in February of 2013 are visualized in figure 4.9. This graph is comprised of around 2,500 ASes and 3,000 different suspicious announcements. An edge between two ASes implies that these ASes, are involved in a suspicious BGP announcements, either as Origin of the prefix, or as Suspicious intermediate AS in the AS-path.

In order to facilitate the analysis of these announcements filtering is employed. Specif-

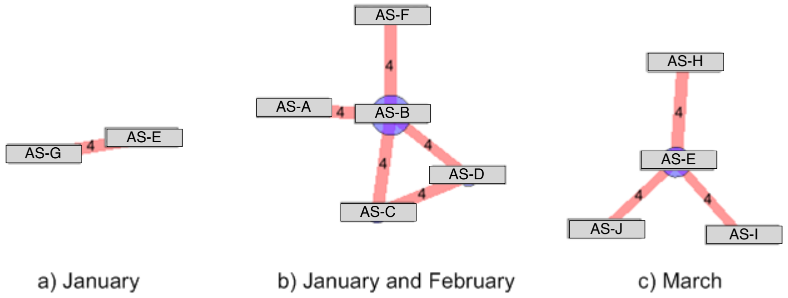


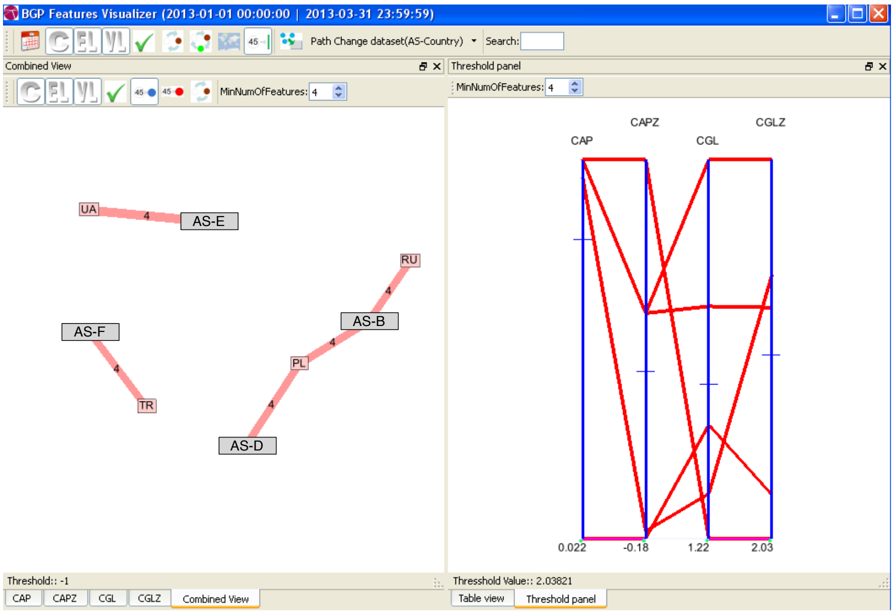
Figure 4.12: The tree groups depicted in figure 4.11 and the corresponding active time periods.

ically, using the parallel coordinates filtering method provided by BGP Features Visualization, the analyst can focus only on a small subset of the most suspicious events for attack detection and attribution. The visualization approach after the application of filtering is depicted in figure 4.10. This figure shows the most interesting cases. Correlating this result with the outcome of the SpamTracer dataset, we indeed discover that the prefixes involved in the formation of a specific group of ASes are also involved in spam activities. The corresponding group is the largest one depicted in figure 4.10. As it can be seen from this figure, this group of connected ASes is comprised of AS-A, AS-B, AS-C, and AS-D. These ASes are either owners of the announced prefixes of suspicious intermediate ASes. Either way, they are involved in suspicious BGP and spam activities that must be further investigated.

This result is not surprising, and implies the existence of an underling relationship between malicious BGP activity and spam activity. In other words, it is highly probable that spammers utilize BGP hijack in an effort to hinder traceability and defeat sender reputation based spam filtering.

4.2.2.2 Checking spam prefixes for suspicious BGP activity

The analysis that will be presented in this Section will investigate prefixes that have known spam activity, for the purpose of correlating this activity with suspicious behavior regarding BGP. For this purpose, two approaches will be combined, the Spamtracer approach so as to provide a list of suspicious prefixes, and the BGP Features Visualizer approach to visualize these prefixes and detect possible malicious activity with regards to BGP. Figure 4.11 shows the visualization of all the ASes that are related to spam prefixes for the period from 01-Jan-2013 to 31-Mar-2013.


BGP Features Visualizer (2013-01-01 00:00:00 | 2013-03-31 23:59:59)

Path Change dataset(AS-Country)

MinNumOfFeatures: 4

Threshold:: -1

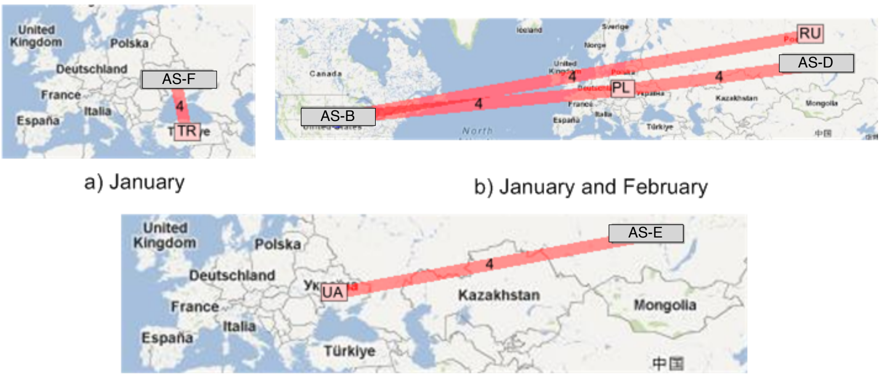
Thresshold Value:: 2.03821

Figure 4.13: The BGP Features Visualizer visualization of all the spam prefixes for the period from 01-Jan-2013 to 31-Mar-2013. The depicted ASes are performing suspicious BGP activity against specific Countries, which host the spam prefixes. This figure shows the combined graph view and Parallel Coordinates User Interface.

This figure depicts either suspicious ASes or owners of the prefixes. Using the BGP Features Visualizer approach it is indeed obvious that all of the spam prefixes exhibit extremely low probability (CAP) and z-score of probability (CAPZ) values, below 0.02 and -0.18 respectively. Furthermore, most of these prefixes also exhibit relatively large values with regards to geographic deviation (CGL) and z-score of geographic deviation (CGLZ). This indicates possible anomalies in BGP, which in turn implies an underling correlation between BGP activity and spammers.

As it can be seen in figure 4.11, the involved ASes form three groups. These groups as well as the activity period of each group are depicted in figure 4.12. As we can see, these groups indicate specific periods of activity.

Changing the view of BGP Features visualizer to AS-Country mode, i.e. show the suspicious ASes and the Countries of origin of the monitored prefixes, figure 4.13 is created. Again as with the previous views, this figure is also comprised of three groups



a) January

b) January and February

c) March

Figure 4.14: The tree groups depicted in figure 4.13 and the corresponding active time periods. The nodes of the graph are positioned in their corresponding postilion on the globe, i.e Country nodes to the corresponding country and AS nodes to their country of origin. The depicted ASes are related to spam prefixes and are also involved in suspicious BGP activity.

of subgraphs.

Closer examination of these groups in figure 4.14 reveals the suspicious ASes as well as their geographic relationship with regards the target countries that host the spam prefixes. The main targets are four countries, namely: Russia, Turkey, Poland, and Ukraine. Specific intermediate ASes toward the paths of each country have high degree of anomaly.

4.2.3 Routing and Spamming Behavior of Fly-By Spammers

In the previous two sections we presented the visual analysis of a few suspicious fly-by spammer cases highlighting the complementarity of different network analytics and visualisation techniques developed in the context of VIS-SENSE. In this section we elaborate on the routing and spamming behavior of fly-by spammers to shed some light on the way those spammers hijack networks and how effective they are at circumventing existing protections, e.g., spam filters.

Figure 4.15 presents the routing and spamming history of nine network address blocks we suspect were hijacked by fly-by spammers. The line in blue represents, for each address block (X-axis), the time the network was announced in BGP. The coloured dots inside the blue lines (from yellow to red) represent, for each address block, the amount

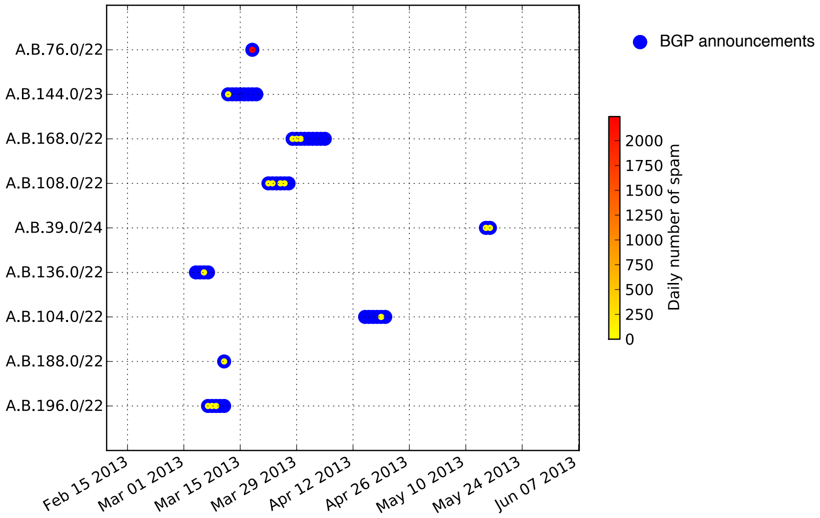


Figure 4.15: Spam received from some networks hijacked by fly-by spammers

of spam received daily from the network. This figure thus highlights (i) the strong *temporal correlation* between BGP announcements of networks and spam (i.e., networks are only announced when spam is received) and (ii) the *short-lived* nature of the BGP announcements (e.g., from one day to eight days on the figure). Moreover spam received from those likely hijacked networks could not be mapped to any known spam botnet which reinforces the idea of those networks being hijacked by fly-by spammers. On the one hand, when a fly-by spammer hijacks a network which is normally unadvertised, he allows hosts he is actually in control of to be connected to the Internet with IP addresses of the hijacked victim network. On the other hand, botnets consist of infected machines (i.e., bots) which are already connected to the network of their victim owner. Thus (spam) bots should not normally be observed in a hijacked network.

In order to assess the effectiveness of suspected fly-by spammers to circumvent anti-spam techniques, we collected the records for the suspicious networks in the Uceprotect spam sender blacklist (DNSBL) [1]. Figure 4.16 presents the routing history of the nine case studies network as well as the number of blacklisted hosts within each address block. We can clearly see that out of the nine networks only two had at least one IP address blacklisted [2]. From this figure it appears that, by sending spam from short-lived hijacked

[1] hbp://www.uceprotect.net/

[2] Blacklist entries automatically expire after 7 days.

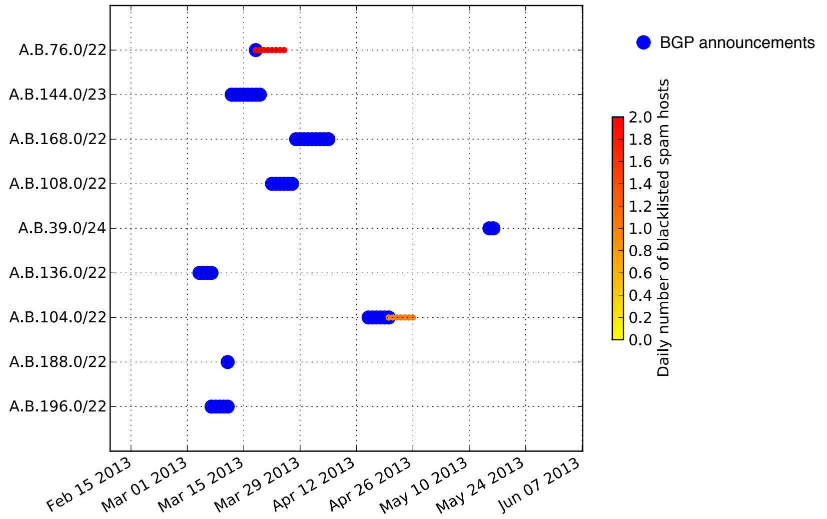


Figure 4.16: Blacklist records for some networks hijacked by fly-by spammers

networks, fly-by spammers are quite successful at circumventing blacklists which many spam filters still heavily relies on as a first layer of defence.

Finally, an analysis of spam campaigns launched from networks hijacked by fly-by spammers carried out using the Triage framework and levering the developed visualisations is presented in Deliverable 6.1 ”Threat Landscape Identification Scenario”. In this analysis we were able to link some spam campaigns launched from different hijacked prefixes indicating that they probably share some common root cause, e.g., launched by the same gang of spammers.

4.2.4 Conslusion

In conclusion, by investigating all the routing information, as well as the spam prefixes, using the different network analytics and visualisation tools a certain degree of correlation was detected between spammers and suspicious BGP activity. This implies that spammers probably exploit BGP vulnerabilities, so as to either hide their identities and ovoid traceability or overcome the IP based filtering applied by spam filters.

4.3 Case Study III: The Bulgarian Case

4.3.1 Introduction

This Section focuses on the analysis of a case that was detected following a correlation of alerts between SpamCloud and the MOAS filtering technique coupled with BGPDB (Section 2.1.6). The events occurred between December 2012 and March 2013, and a MOAS conflict was detected on February 3rd and can be summarizer as follows. In the original situation, the prefix owners announces an aggregated class B prefix, and has been doing so for many years. Then, in December, a new entity starts announcing a set of subprefixes of the owner. In February, another entity starts announcing those subprefixes, resulting in a MOAS conflict, which we successfully detected. Eventually, the initial situation resumes when all subprefixes are withdrawn.

This case shows just how difficult it is to validate a hijack. We have a strong correlation between a routing anomaly and spam, traffic information confirming the spam and scam activities. This is already more information than Ramachandran et al. [34] used in order to conclude the existence of fly-by spammers. However, we dug deeper in order to find authoritative data that would confirm the hijack. *These sources eventually revealed that the IP subspace was rented out from the owner and abused. The interest in this case, then, lies in the demonstration of the critical need to validate hijackings: even though the arguments for hijacks are very strong – previous work would have classified this event as a malicious hijack – there was no hijacking. Consequently, this case demonstrates the limit of certitude that can be reached by using publicly available data.*

4.3.2 Evidences For The Hijack

For the month of February 2013, 2,331 distinct prefixes were involved in control plane alerts, i.e., MOAS conflicts. We use a time window of 15 days to correlate these events with spam from IP addresses observed at spam traps and on blacklists to identify malicious hijacks. In the following we present an in-depth analysis for one of the matching events. Note that all results are anonymized with good cause; we are nevertheless willing to share details upon request.

Based on several alarms raised by our detection system on February 3rd, 2013, we became aware of an incident taking place in Bulgaria. Several MOAS conflicts were observed for networks that correlated with emerging spamming activities. We carried out a detailed analysis of these events, and present our results in chronological order below.

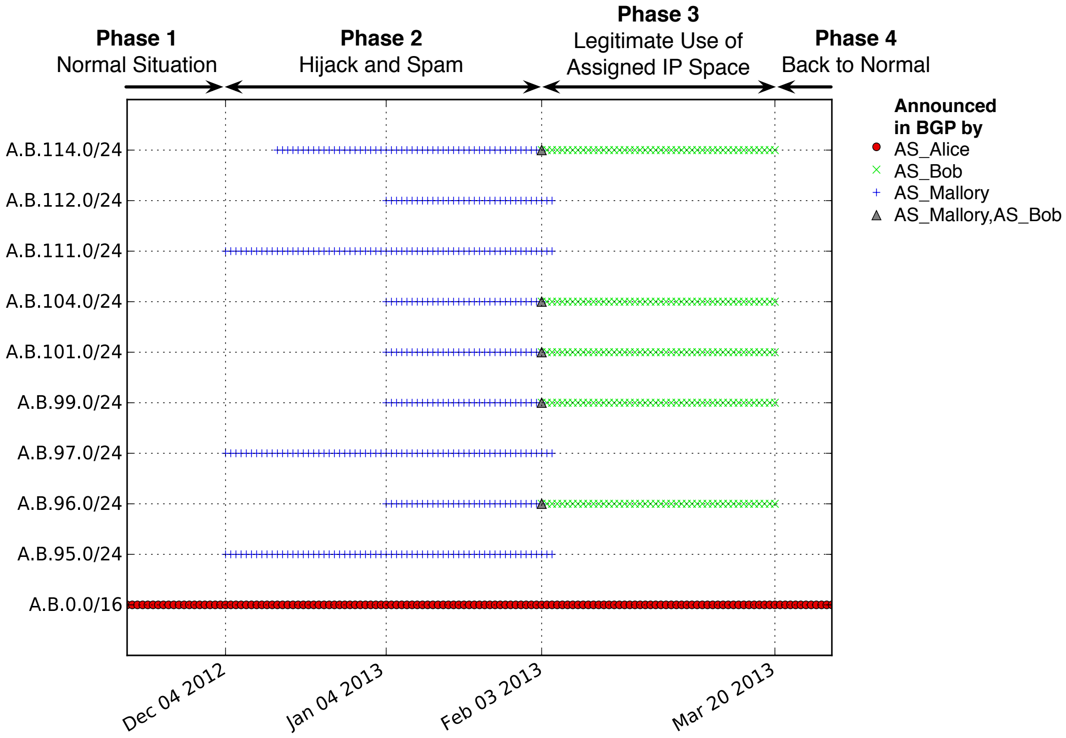


Figure 4.17: Route announcements for the *Bulgarian Case*

Phase 1: Normal Situation

Since 2008, the prefix `A.B.0.0/16` has been announced in BGP by a Tier-3 ISP *Alice*. This ISP is known to provide hosting services for a variety of customers. We did not observe announcements of more specific prefixes during the whole time of phase 1 (Figure 4.17).

Phase 2: Hijack and Spam

On December 4, 2012, *Mallory* started announcing a set of nine more specific (`/24`) prefixes of *Alice*, who carried on with the original `/16` announcement (Figure 4.17). By using online *WHOIS* queries, we learned that *Mallory* supposedly is a VPS service provider also located in Bulgaria. A thorough web-search however returned no result for this specific company.

Spam

Figure 4.18 shows spam[3] received by Symantec.cloud spam traps from IP addresses belonging to the nine prefixes announced by *Mallory*. Figure 4.19 presents blacklisted IP addresses from Uceprotect Level-1 [6] related to these prefixes. This figure shows a strong correlation between the BGP routing announcements, spam, and blacklisted IP addresses. On some days, up to 80 spam emails were sent to our spam honeypots. Many prefixes also had around 100 backlisted IP addresses for several days. On Figure 4.19 we still observe some blacklisted IP addresses after the end of phase 2 but we attribute them to the one-week expiration period of Uceprotect records. Symantec.cloud spam dataset may provide the spam botnet name responsible for the spam based on spam bot signatures. Because spam bots are usually compromised machines, they should not be observed on hijacked IP space. And indeed no such botnet could be inferred from Symantec.cloud's reports for spam hosts in the suspicious prefixes. This indicates that those machines were likely set up by the spammers themselves.

Scam Hosting Infrastructures

We further analyzed the spam mails and were able to identify several URLs within these messages. Out of 118 extracted domain names, 89 resolved to an IP address within six of the obtrusive prefixes. We conclude that the spam was also used as a platform to promote a scam infrastructure hosted within these prefixes. About 90% of all scam hosts in the nine `A.B.x.0/24` networks coincided with IP addresses of spam hosts, which indicates that the spammers took full advantage of the prefixes under their control.

[3] Live spam feed of ∼4M spam per day starting in January 2012

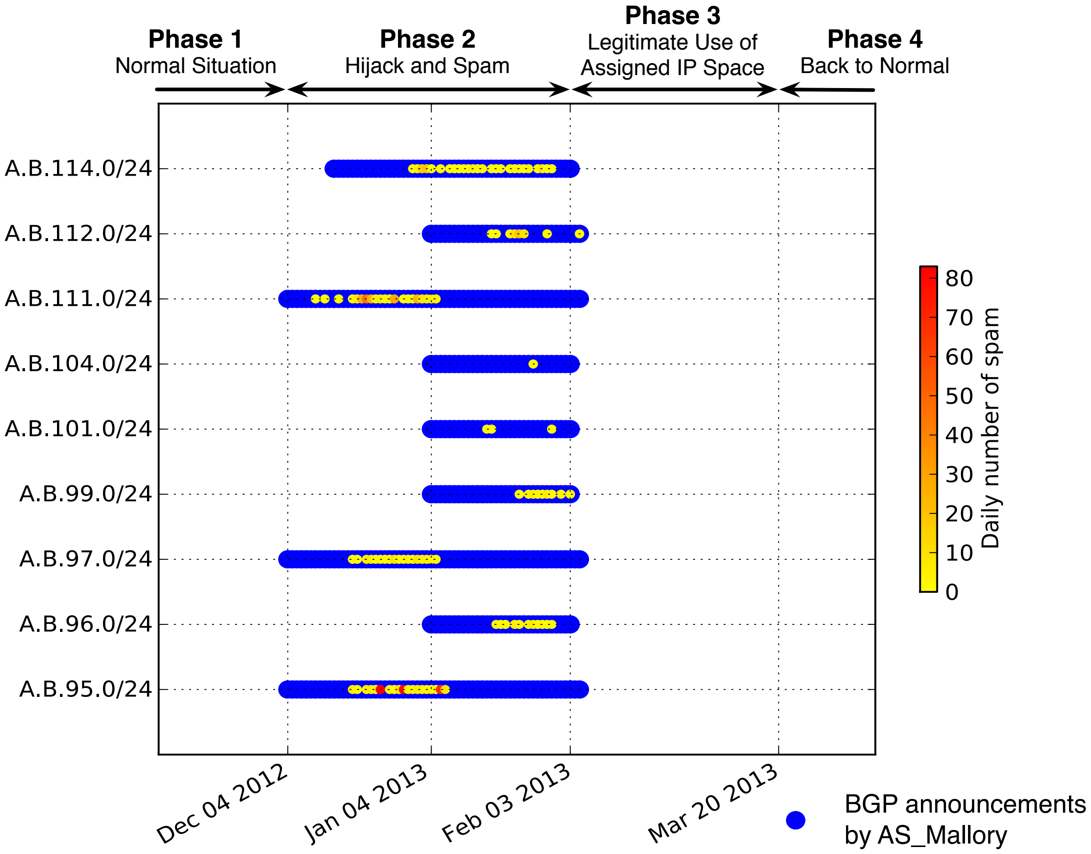


Figure 4.18: Spam received from reported prefixes

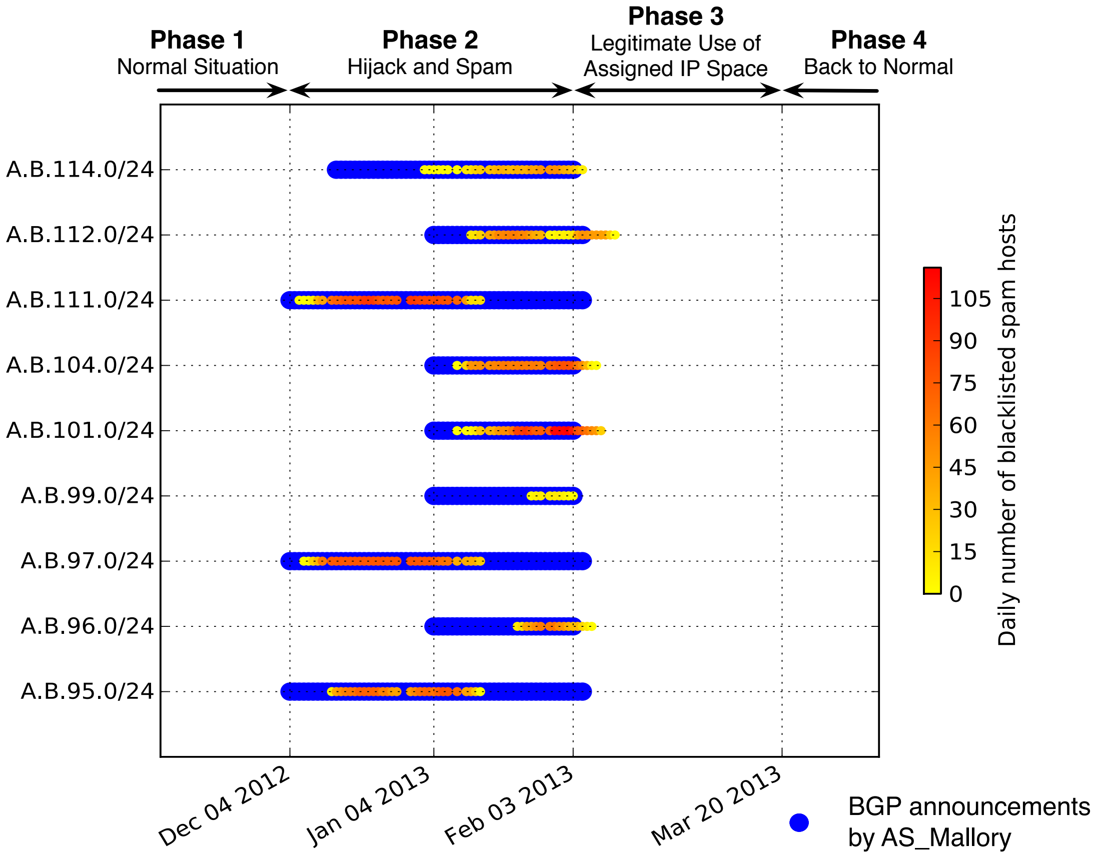


Figure 4.19: Blacklist records for reported prefixes

It is interesting to see that almost all scam hosts' IP addresses shared the same last byte while being spread over all abused networks (e.g. A.B.{95,96,114}.**5**, A.B.{95,114}.**9**, A.B.{95,96,97,114}.**14**, etc). Similar characteristics appear for the resolution of domain names to IP addresses within the nine prefixes. All 89 resolvable domains were created at nearly the same time as the prefixes were first announced in BGP by *Mallory*. All pieces of evidence suggest a single administrator behind the domains and network infrastructure.

Netflow Traffic Analysis

In addition, we look at netflow data to analyze changes in traffic patterns before, during and after a suspected hijack. Such changes can range from simple outages in monitored networks, where outgoing connection attempts are unanswered, to changes in traffic volume or even to a significant amount of new connections from and to different sets of ports. We utilize archived netflow data of the Münchner Wissenschaftsnetz (MWN) – Munich's scientific network – which comprises more than 80,000 end hosts. It is used by researchers, students, and administrative personnel, who generate monthly upstream and downstream traffic volumes of more than 300 and 600 Terabyte, respectively. We consider the MWN large enough to be effected by large-scale spam campaigns, and expect to observe at least some portions of spam that originate from hijacked networks.

We were able to isolate 13k inbound flows from the suspicious prefixes for the period of December 2012 to March 2013. The majority of these flows accounted for SMTP requests (71.0%), DNS replies (25.2%), HTTP replies (1.6%) and SMTP replies (1.4%). The remaining 1.8% of flows indicated traffic to an IRC server within our networks, and to ephemeral UDP ports. For 97.4% of all incoming flows, we observed corresponding outgoing flows. An analysis of the IRC traffic revealed that these flows originated from 1,381 hosts spread over 254 different `/24` subnets within the `/16` prefix announced by *Alice*. Such orchestrated IRC traffic across all networks of *Alice's* customers seems to be implausible: we thus assume that these flows attribute to IP spoofing activities unrelated to the Bulgarian case, and exclude them from our analysis.

All connection requests (incoming for SMTP and outgoing for DNS and HTTP) are depicted in Figure 4.20. We observe a strong correlation in phase 2 between the BGP announcements (Figure 4.17), the observed spam (Figure 4.18), and the blacklist records (Figure 4.19). We observed a total of 925 IP addresses for the delinquent's activities, of which 850 IP addresses were used to send spam mail. Less than 10% of these addresses were re-used for the DNS and HTTP activities. We further found 30 distinct DNS servers mostly hosted in the prefixes `A.B.96.0/24` and `A.B.114.0/24`, which were queried over 3,000 times by clients in our networks. The flow data also shows 200 bidirectional HTTP connections to more than 100 web servers in the reported prefixes.

This analysis confirms that the prefixes were used in order to massively send spam from several hundred clients. Furthermore, it clearly shows that the person in charge hosted more than 100 live services (DNS and HTTP), presumably to do phishing or similar fraudulent activities.

Phase 3: Legitimate Use of Assigned IP Space

On February 3, 2013, *Bob* started announcing five of the nine prefixes announced by *Mallory*, resulting in MOAS conflicts during a few hours before *Mallory* withdrew all of its announcements. *Alice*, once more, kept on announcing the original `/16` prefix (Figure 4.17). Several spam hosts that used to reply to traceroute probes on consecutive days during phase 2 also suddenly became unreachable suggesting a real change in network topology.

Bob is a business-to-business IT service provider located in the same country as *Mallory* and *Alice*, according to their website. Its ASN first appeared in BGP in November 2008. All five `/24` prefixes were announced via *Alice* acting as legitimate upstream provider. Figure 4.21 depicts the overall topology from a BGP's point of view.

With beginning of phase 3, all malicious activities suddenly stopped. This indicates that *Bob* was regularly assigned the five prefixes by *Alice* in the context of a provider-to-customer business relationship.

Phase 4: Back to Normal

On March 20, 2013, *Bob* withdrew its announcements of *Alice's* five prefixes, resulting in the same initial situation as described for phase 1, where the whole prefix `A.B.0.0/16` was announced by *Alice* only.

Given these findings, approaches presented in [34, 22] would conclude the existence of a malicious BGP hijack. All evidence presented so far, especially the strong correlation for both the control plane and the data plane, lead us to the conclusion that we indeed observed a malicious hijacking event for this *Bulgarian Case*.

4.3.3 Evidences Against the Hijack

Despite the evidence for a malicious hijack incident described so far, we decided to further investigate the case and found significant evidence *against a hijacking event*. We analyzed more than one year of archived RIPE IRR database dumps in order to infer the legitimate owners of the suspected prefixes by searching for `route` objects and looking into the corresponding *origin (AS)* attributes. We found that *Alice* carefully

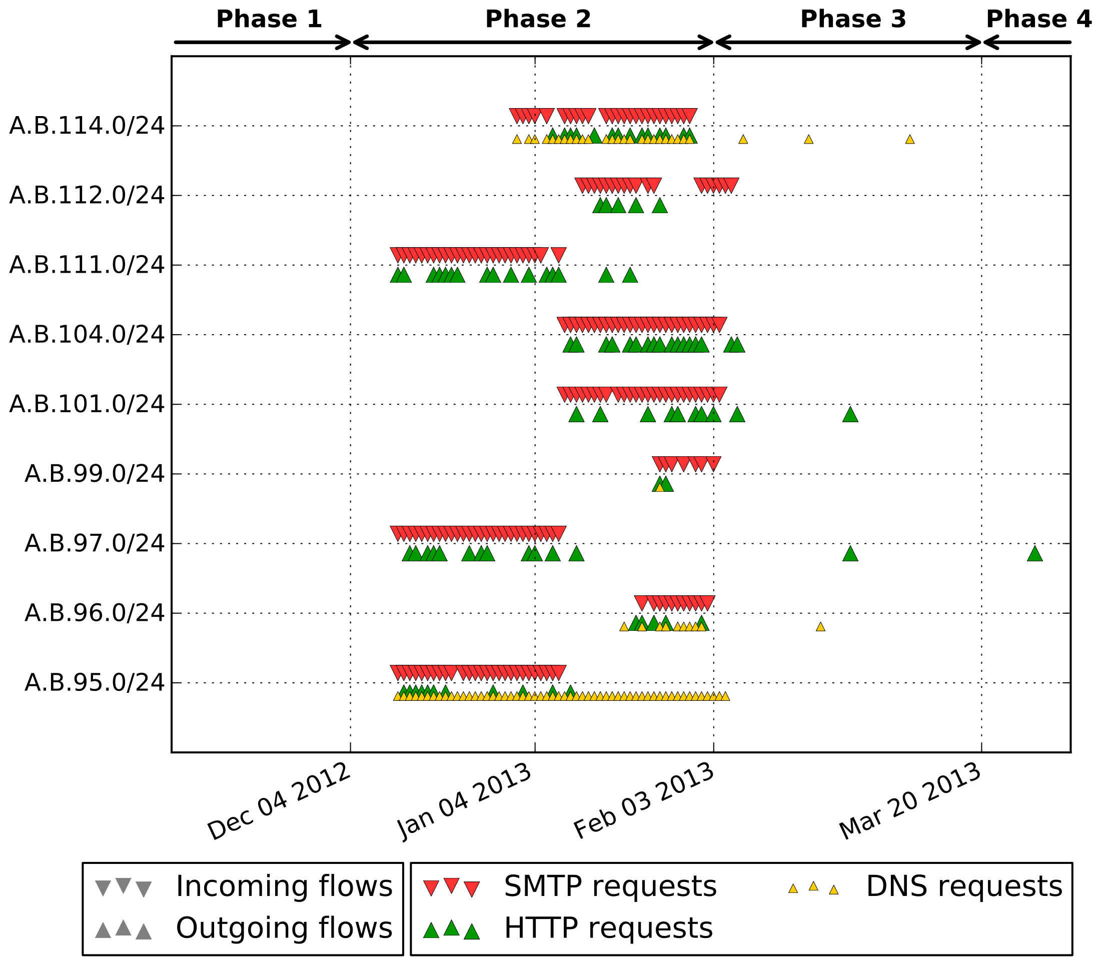


Figure 4.20: Flow data for reported prefixes

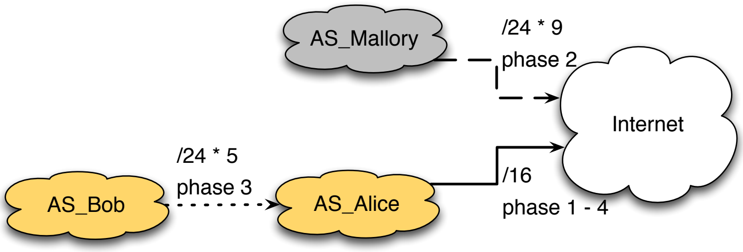


Figure 4.21: Topology derived from BGP

maintained such `route` objects in the RIPE IRR database throughout all four phases. We obtained the first three objects related to the prefixes in question on December 4th, 2012 (Figure 4.22). Their *origin* attributes were set to *Mallory*, and the creation time corresponded to her first BGP announcements. This clearly indicates that – at least according to the RIPE IRR database – *Mallory* was authorized to use these prefixes.

Figure 4.22 gives an overview for all relevant `route` objects that we found in the RIPE IRR database. We learned that the dates of appearance fully match all BGP announcements of *Mallory* and *Bob* (see Figure 4.17), and all objects were maintained by *Alice*. If we assume that an attacker is incapable to alter the RIPE IRR database at will (and that he had no access to *Alice's* maintainer account), we must conclude that *Alice* delegated all nine prefixes to *Mallory* by choice, and reassigned some of them around February 3rd, 2013 to *Bob*.

We further extracted the database objects' *descr* attributes, and even found some weak evidence for a relationship between *Mallory* and *Bob*. Those free text fields can be set to any value. For *Mallory*, all fields were set to `BG-XX-N`. `BG` indicates Bulgaria, whereas `N` corresponds to each of the prefixes' third byte. More importantly, `XX` represented the initial letters of *Bob*'s company name. After reassignment, the description changed to *Bob*'s full company name.

Finally, we contacted *Mallory's* upstream provider and learned that Mallory requested to announce rented prefixes. After receiving complaints, the upstream provider cancelled *Mallory's* contract.

Given all circumstances, we must conclude that *Mallory* acted maliciously by sending spam. However, we cannot decide if Mallory really hijacked prefixes, or if Mallory just rented the networks for abuse.

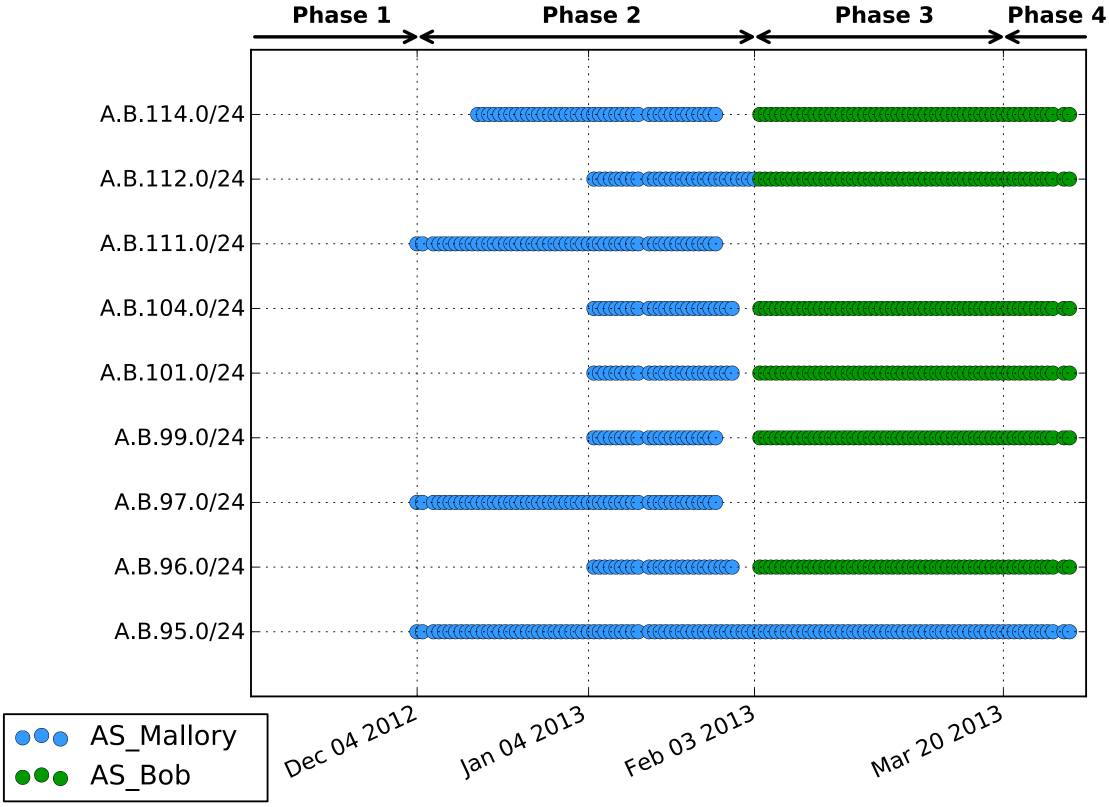


Figure 4.22: RIPE IRR `route` objects for reported prefixes

4.3.4 Discussion

Even though we have accumulated a series of converging clues incriminating one of the actors, namely *Mallory*, involved in performing BGP hijacks with malicious intent, we still cannot reach a decisive conclusion.

As presented in Section 4.3.2, the strong correlation between the BGP announcements of *Alice's* sub-prefixes by *Mallory*, the spam received by Symantec.cloud and the evidence of scam hosting infrastructures during phase 2 initially led us to believe that *Mallory* had indeed hijacked these prefixes to emit spam. This result is supported by the following observations:

1. The temporal correlation between the BGP announcements and the emerging spam during phase 2 strongly suggests that machines in *Mallory's* network are the spam sources.

2. *Mallory*'s first appearance in BGP as well as the registration date of the domain names advertised in the received spam mails directly coincident with phase 2 of the incident.

3. *Alice* provided upstream connectivity for *Bob*, while *Mallory* hired an independent upstream provider, although *Alice* continuously announced the full enclosing `/16` prefix.

4. As soon as *Bob* started to announce his assigned prefixes in phase 3, *Mallory's* announcements and the emission of spam stopped, and no more traffic flows were observed.

Our findings in Section 4.3.3 validate prefix ownership based on the RIPE IRR database for all involved parties during all phases of the incident. However, this fact does not exclude a malicious BGP hijack: it is possible that an attacker covered up his traces by altering objects in the RIPE IRR database. According to RIPE, 86% of database maintainers were using password-only authentication in 2011 [39]. However, password protection may not be enough since an attacker could use information leaked from the IRR database [10] and/or phishing e-mails [8] to gain privileged access to the database.

Our system to detect malicious BGP hijacks was partly designed upon findings of previous studies on the root causes of BGP hijack events, like Ramachandran et al.'s study [34] on short-lived BGP announcements, the correlation between BGP hijack alerts and spam by Hu et al. [22] and a validated hijack case performed by a spammer described by Vervier et al. in [48] and by Schlamp et al. in [41]. Comparing our findings presented in Section 4.3.2 with those reported in previous work quickly led us to the conclusion that the *Bulgarian Case* was indeed a malicious BGP hijack. However, the novel forensic analysis of an IRR database described in Section 4.3.3 at least opened our mind that we possibly have not found a real hijack event, but rather a plain abuse of rented IP space. *In the end, although we remain indecisive, we learned that it is crucial to consider complementary data sources, preferably as independent as possible (e.g. IRRs) as well as feedback from network owners (e.g. via mailing lists like NANOG as in [48, 41]) in order to avoid drawing conclusions too quickly based on a limited set of evidence skewed toward one verdict or the other.* This fact is of particular interest to avoid mis-attributing attacks launched from hijacked IP space when responding with possibly legal actions.

5 Summary and Conclusion

In this document, we took advantage of the BGP perspective of the VIS-SENSE framework in order to investigate malicious BGP hijacks.

Chapter 2 introduced the analytical methods and visualizations that were designed during WP3 and WP4 for BGP.

Chapter 3 discussed the definition of prefix hijacking applied during WP6. This definition differs from the one usually used in the literature because it explicitly considers only cases where the routing infrastructure was attacked in order to carry out malicious activities on the Internet. This means that any non-malicious hijack, resulting from, for example, router misconfiguration, operational errors, are not part of this analysis. The severe lack of access to ground-truth data was addressed, and solutions were presented in order to circumvent the problem. Additionally, weaknesses embedded in BGP data due to protocol designed were discussed.

Chapter 4 presents validated hijack cases. The Link-Telecom hijack is a case where an American spammer abused a Russian ISP's abandoned prefix during four months in 2011. The case is analyzed with the help of the visualization tools provided by the VIS-SENSE framework. The fly-by spammer cases analyzes a set of dormant prefixes that have been announced in order to emit spam during the first six months of 2013. These cases were uncovered by the use of BGP Features Visualizer on prefixes monitored by SpamTracer. Finally, the Bulgarian case shows how critical validating hijacks is by looking at a case where a spammer appears to have abused a Bulgarian ISP's subprefixes in order to carry out spam and scam activities. A thorough investigation eventually rules out the possibility of a hijack. However, weaker analysis methods – consistently used by previous research on the topic – would have concluded on a hijack occurrence.

This Deliverable illustrates the capabilities of the VIS-SENSE framework which leverages a diverse, rich set of data: security data (SpamCloud), as well as network data (BGPDB, SpamTracer). By taking advantage of the WAPI servers associated with these data sources, data correlation can easily be carried out in order to eliminate as many benign events as possible. Visualization tools can then be used on top of this correlated output in order to further filter out benign events, and also to isolated events worth of being further analyzed.

Bibliography

[1] Composite blocking list (dnsbl). `http://cbl.abuseat.org`.

[2] Inferred AS Relationships. `http://as-rank.caida.org/data/`.

[3] North American Network Operators' Group. `http://www.nanog.org/`.

[4] Ripe network coordination centre). `http://www.ripe.net/`.

[5] Ripe ris. `http://www.ris.ripe.net`.

[6] Uceprotect. `http://www.uceprotect.net`.

[7] Prefix hijacking by Michael Lindsay via Internap. `http://mailman.nanog.org/pipermail/nanog/2011-August/039381.html`, August 2011.

[8] Dear RIPE: Please don't encourage phishing. `http://mailman.nanog.org/pipermail/nanog/2012-February/045062.html`, February 2012.

[9] Symantec Internet Security Threat Report: Future Spam Trends: BGP Hijacking. Case Study - Beware of "Fly-by Spammers". `http://www.symantec.com/threatreport/`, April 2012.

[10] Whois.afrinic.net leaks passwords. `https://lists.afrinic.net/pipermail/rpd/2012/002586.html`, November 2012.

[11] H. Ballani, P. Francis, and X. Zhang. A Study of Prefix Hijacking and Interception in the Internet. In *SIGCOMM '07: Proceedings of the 2007 conference on Applications, technologies, architectures, and protocols for computer communications*, pages 265–276, New York, NY, USA, 2007. ACM.

[12] E. W. Biersack, Q. Jacquemart, F. Fischer, J. Fuchs, O. Thonnard, G. Theodoridis, D. Tzovaras, and P.-A. Vervier. Visual analytics for BGP monitoring and prefix hijacking identification. *IEEE Network Magazine, Special Issue on Computer Network Visualization, November/December 2012*, 11 2012.

[13] V. J. Bono. 7007 explanation and apology. `http://www.merit.edu/mail.archives/nanog/1997-04/msg00444.html` (Retrieved on November 23rd, 2010), April 1997.

[14] K. Butler, T. Farley, P. McDaniel, and J. Rexford. A survey of bgp security issues and solutions. In *Proceedings of the IEEE*, volume 98, pages 100–122, January 2010.

[15] K.-W. Chin. On the characteristics of BGP multiple origin AS conflicts. In *Telecommunication Networks and Applications Conference, 2007. ATNAC 2007. Australasian*, pages 157–162, 2007.

[16] R. Dashboard. Distribution of RPKI states. `http://rpki.surfnet.nl/global.html` (Retrieved on October 3rd, 2013).

[17] F. Fischer, J. Fuchs, P.-A. Vervier, F. Mansmann, and O. Thonnard. VisTracer: A Visual Analytics Tool to Investigate Routing Anomalies in Traceroutes. In *To appear in Proc. of the 9th International Symposium on Visualization for Cyber Security (VizSec 2012)*. ACM Digital Library, 2012.

[18] A. Freedman. 7007: From the horse's mouth. `http://merit.edu/mail.archives/nanog/1997-04/msg00380.html` (Retrieved on November 23rd, 2010), April 1997.

[19] L. G. L. Gao. On inferring autonomous system relationships in the Internet. *IEEE/ACM Transactions on Networking*, 9(6):733–745, 2001.

[20] J. Hawkinson and T. Bates. Guidelines for creation, selection, and registration of an Autonomous System (AS). RFC 1930 (Best Current Practice), March 1996.

[21] S.-C. Hong, H.-T. Ju, and J. W. Hong. IP prefix hijacking detection using idle scan. In *APNOMS'09: Proceedings of the 12th Asia-Pacific network operations and management conference on Management enabling the future internet for changing business and new computing services*, pages 395–404, Berlin, Heidelberg, 2009. Springer-Verlag.

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

[23] J. Karlin, S. Forrest, and J. Rexford. Pretty good bgp: Improving bgp by cautiously adopting routes. In *Proceedings of the IEEE International Conference on Network Protocols 2006 (ICNP '06)*, pages 290–299, Santa Barbara, CA, USA, Nov. 2006.

[24] Khare, V. and Ju, Q. and Zhang, B. Concurrent Prefix Hijacks: Occurrence and Impacts. In *IMC*, pages 29–36. ACM, 2012.

[25] C. Labovitz. Additional discussion of the april china bgp hijack incident. http://asert.arbornetworks.com/2010/11/additional-discussion-of-the-april-china-bgp-hijack-incident/ (Retrieved on November 23rd, 2010), April 2010.

[26] C. Labovitz. China hijacks 15% of internet traffic? http://asert.arbornetworks.com/2010/11/china-hijacks-15-of-internet-traffic/ (Retrieved on November 23rd, 2010), April 2010.

[27] M. Lad, D. Massey, D. Pei, Y. Wu, B. Zhang, and L. Zhang. PHAS: A Prefix Hijack Alert System. In *USENIX-SS'06: Proceedings of the 15th conference on USENIX Security Symposium*, Berkeley, CA, USA, 2006. USENIX Association.

[28] R. Mahajan, D. Wetherall, and T. Anderson. Understanding bgp misconfiguration. *SIGCOMM Comput. Commun. Rev.*, 32(4):3–16, Aug. 2002.

[29] R. McMillan. A chinese isp momentarily hijacks the internet. http://www.nytimes.com/external/idg/2010/04/08/08idg-a-chinese-isp-momentarily-hijacks-the-internet-33717.html (Retrieved on November 23rd, 2010), April 2010.

[30] R. NCC. Routing Information Service, Raw Data. http://www.ripe.net/data-tools/stats/ris/ris-raw-data. [Online; accessed 29-Mar-2012].

[31] C. News. Router glitch cuts net access. http://news.cnet.com/2100-1033-279235.html (Retrieved on November 23rd, 2010), April 1997.

[32] A. Pilosov and T. Kapela. Stealing the internet: An internet-scale man in the middle attack, August 2008. Presentation at DEFCON16, http://www.defcon.org/images/defcon-16/dc16-presentations/defcon-16-pilosov-kapela.pdf (Retrieved on November 16th, 2010).

[33] J. Qiu, L. Gao, S. Ranjan, and A. Nucci. Detecting bogus BGP route information: Going beyond prefix hijacking. In *Security and Privacy in Communications Networks and the Workshops, 2007. SecureComm 2007.*, pages 381–390, 2007.

[34] 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.

[35] Y. Rekhter, T. Li, and S. Hares. A Border Gateway Protocol 4 (BGP-4). RFC 4271 (Draft Standard), Jan. 2006. Updated by RFCs 6286, 6608, 6793.

[36] RIPE. Youtube hijacking: A ripe ncc ris case study. `http://www.ripe.net/news/study-youtube-hijacking.html` (Retrieved on November 23rd, 2010), 2008. RIPE.

[37] RIPE. Youtube hijacking: A ripe ncc ris case study. `http://www.ripe.net/news/study-youtube-hijacking.html`, 2008. RIPE.

[38] RIPE NCC. Bgp origin validation. `http://www.ripe.net/lir-services/resource-management/certification/bgp-origin-validation` (Retrieved on October 2nd, 2013).

[39] RIPE NCC. Authentication Methods Used in the RIPE Database. https://labs.ripe.net/Members/kranjbar/authentication-methods-used-in-the-ripe-database, August 2011.

[40] M. Roughan, W. Willinger, O. Maennel, D. Perouli, and R. Bush. 10 lessons from 10 years of measuring and modeling the internet's autonomous systems. *Selected Areas in Communications, IEEE Journal on*, 29(9):1810–1821, 2011.

[41] Schlamp, J. and Carle, G. and Biersack, E. W. A forensic case study on AS hijacking: the attacker's perspective. *SIGCOMM CCR*, pages 5–12, 2013.

[42] X. Shi, Y. Xiang, Z. Wang, X. Yin, and J. Wu. Detecting prefix hijackings in the internet with argus. In *Internet Measurement Conference*, pages 15–28, 2012.

[43] 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.

[44] G. Siganos and M. Faloutsos. BGP routing: A study at large time scale. In *in Proc. IEEE Global Internet*, 2002.

[45] M. Tahara, N. Tateishi, T. Oimatsu, and S. Majima. A Method to Detect Prefix Hijacking by Using Ping Tests. In *Proceedings of the 11th Asia-Pacific Symposium on Network Operations and Management (APNOMS '08)*, Berlin, Germany, 2008.

[46] A. Toonk. Securing BGP routing with RPKI and ROA's. `http://www.bgpmon.net/securing-bgp-routing-with-rpki-and-roas/` (Retrieved on October 2nd, 2013).

[47] A. Toonk. Chinese isp hijacks the internet. `http://bgpmon.net/blog/?p=282` (Retrieved on November 23rd, 2010), April 2010.

[48] P.-A. Vervier and O. Thonnard. SpamTracer: How Stealthy Are Spammers? In *The 5th IEEE International Traffic Monitoring and Analysis Workshop (TMA)*, Turin, Italy, Apr. 2013.

[49] J. Wolf. Pentagon says "aware" of china internet rerouting. `http://www.reuters.com/article/idUSTRE6AI4HJ20101119?pageNumber=1` (Retrieved on November 23rd, 2010), November 2010.

[50] Z. Zhang, Y. Zhang, Y. C. Hu, Z. M. Mao, and R. Bush. iSPY: Detecting IP Prefix Hijacking on My Own. In *Proceedings of the ACM SIGCOMM 2008 Conference on Data communication (SIGCOMM '08)*, pages 327–338, New York, NY, USA, Aug 2008.

[51] X. Zhao, D. Pei, L. Wang, D. Massey, A. Mankin, S. F. Wu, , and L. Zhang. An analysis of BGP multiple origin AS (MOAS) conflicts. In *Proceedings of the ACM SIGCOMM 2001 Workshop on Internet Measurement(IMW '01)*, pages 31–35, San Francisco, CA, USA, Nov. 2001.

[52] C. Zheng, L. Ji, D. Pei, J. Wang, and P. Francis. A Light-Weight Distributed Scheme for Detecting IP Prefix Hijacks in Real-Time. *ACM SIGCOMM - Computer Communication Review*, 37(4):277–288, 2007.

[53] E. Ziegel, W. Press, B. Flannery, S. Teukolsky, and W. Vetterling. *Numerical Recipes in C: The Art of Scientific Computing*, volume 29. Cambridge University Press, Nov. 1987.