Mind Your Blocks: On the Stealthiness of Malicious BGP Hijacks

Pierre-Antoine Vervier
Eurecom
vervier@eurecom.fr

Olivier Thonnard
Symantec Research Labs
Olivier_Thonnard@symantec.com

Marc Dacier
Qatar Computing Research Institute
mdacier@qf.org.qa

***Abstract*—Some recent research presented evidence of blocks of IP addresses being stolen by BGP hijackers to launch spam campaigns [35]. This was the first time BGP hijacks were seen in the wild. Since then, only a very few anecdotal cases have been reported as if hackers were not interested in running these attacks. However, it is a common belief among network operators and ISPs that these attacks could be taking place but, so far, no one has produced evidence to back up that claim. In this paper, we analyse 18 months of data collected by an infrastructure specifically built to answer that question: are intentional stealthy BGP hijacks routinely taking place in the Internet? The identification of what we believe to be more than 2,000 malicious hijacks leads to a positive answer. The lack of ground truth is, of course, a problem but we managed to get confirmation of some of our findings thanks to an ISP unwittingly involved in hijack cases we have spotted. This paper aims at being an eye opener for the community by shedding some light on this undocumented threat. We also hope that it will spur new research to understand why these hijacks are taking place and how they can be mitigated. Depending on how BGP attacks are carried out, they can be very disruptive for the whole Internet and should be looked at very closely. As of today, as much as 20% of the whole IPv4 address space is currently allocated but not publicly announced, which makes it potentially vulnerable to such malicious BGP hijacks.**

I. Introduction

The current Internet routing infrastructure is known to be vulnerable to BGP hijacking which consists in taking control of blocks of IP addresses without any consent of the legitimate owners. This is due to the fact that BGP [36], the de facto inter-domain routing protocol, relies on the concept of trust among interconnected autonomous systems (ASes). Accidental, not necessarily malicious, BGP hijack incidents are known to occur in the Internet. They are generally attributed to misconfigurations. A few cases have received public disclosure on network operational mailing lists, such as NANOG, or blog posts [12], [15], [16], [21]. Techniques to detect these BGP hijacks have been proposed to help network operators monitor their own prefixes to react quickly to such possible outages. These approaches suffer from a very high false-positive rate [23], [29], [40], [50] which is still acceptable to these users since they are only interested in alerts related to the networks they own. Other proposals aim at preventing BGP hijacks [24], [25], [30] but their large-scale adoption and deployment are hindered by the implementation cost.

In 2006, Ramachandran et al. [35] introduced a new phenomenon called "BGP spectrum agility", which consists of spammers advertising for a short period of time (*i.e.,* less than one day) BGP routes to large (*i.e.,* /8) previously unannounced blocks of IP addresses and, subsequently, using the available IP addresses for spamming. Later, some other authors also identified the emission of spam emails coming from hijacked prefixes [23], [20]. Furthermore, complementing the work done in [39], we have described in [47] a special case of hijack in which a couple of IP address blocks were stolen and used to send spam. Most recently, we have also shown in [46], thanks to another real-world case, that correlating routing anomalies with malicious traffic, such as spam, is not sufficient to decisively prove the existence of a malicious BGP hijack.

Besides these sparse cases and despite the apparent desire of some owners to detect whether their own IP address block could ever be stolen, to the best of our knowledge we have no documented evidence that BGP attacks are a threat worth being investigated, since no one has shown that hackers have the possibility to routinely use that modus operandi to commit nefarious activities. If they were capable of it, this would constitute a very serious threat to the Internet since this would enable them not only to send spam emails while defeating the classical IP blacklists but, more importantly, to run large scale DDoS at almost no cost or run man-in-the-middle attacks against almost any target of their choosing.

Therefore, we feel that there is a need to rigorously assess the existence and prevalence of this potential threat. This paper contributes to this objective by providing an in-depth study of a specific class of agile spammers that are able to hijack routinely, persistently and -quite likely- automatically a large number of blocks of IP addresses to send spam from the stolen IP space. What truly separates our work from others is that we do not only uncover a large number of new suspicious cases but we also gather contextual information to build a compelling case for the reality of these malicious hijacks and to expose the modus operandi of the malicious actors. Our results show that the identified hijacks were rather successful at circumventing traditional BGP hijack and spam protection techniques.

The main contribution of this paper is to be an eye-opener to the fact that frequent, persistent and stealthy BGP

NDSS'15, 8-11 February 2015, San Diego, CA, USA

http://dx.doi.org/10.14722/ndss.2015.23035

hijack attacks have taken place in the Internet for months or even years. We achieve this by leveraging a large-scale data collection system that we introduced in [47]. This large dataset is then mined by means of a multi-stage scoring and filtering process, whose results are enriched using external data sources and clustered in order to reveal the modus operandi of the attackers.

Before moving into the core of the paper, we would like to make it clear what this paper is *not* about:

- This paper does not offer yet another BGP hijack detection technique. We do not want to incrementally improve the state of the art in that space. The novelty, at the contrary, resides in the identified hijacks and in their detailed analysis.
- Along the previous point, we do not claim that we have found all BGP hijacks that could have been found in our dataset. What matters is that our results must be seen as a proof of the existence of these recurring attacks. Some of them were confirmed by an ISP who was unwittingly involved in several hijack cases apparently performed by a Russian spammer. We believe that others are still hidden in our dataset. However, figuring out how to find them all is left as an exercise for future work.
- This paper does not discuss the optimal choice of the few parameters used in our algorithms. Running a rigorous sensitivity analysis to further improve our results is something we are working on. However, the key contribution of this paper is not the method itself but, instead, the identification of real malicious attacks routinely happening in the wild.

The rest of this paper is organised as follows. Section 2 presents the related work. Section 3 describes the environmental setup we have built, namely the data collection and analysis processes. Section 4 is the core of the paper and goes into the details of all the results obtained when using this environment for several months. Section 5 offers some insights on the effectiveness of current counter-measures to defeat the attacks we have found. Section 6 summarizes the lessons learned and concludes the paper.

II. Related Work

BGP hijacking defense solutions are twofold: (i) detection techniques aim at monitoring the Internet routing infrastructure and trigger alarms upon abnormal routing changes, and (ii) other techniques aim at providing BGP with new mechanisms to mitigate or prevent hijacking. Unlike hijack mitigation or prevention techniques, detection methods require no changes to router software, which usually makes them readily and easily deployable.

Some techniques have been proposed to bring security into BGP [24], [25], [30], usually using cryptography to sign some elements of BGP updates to ensure routing information authenticity and integrity. In the last few years a BGP security framework relying on a RPKI [31] to secure IP prefix origination [24] has gained a lot of attention and is now progressively being deployed.

Alternatively, some existing proposals [29], [26], [28], [34] aim at detecting IP prefix hijacking by passively monitoring the routing infrastructure. However due to the strong similarity between IP prefix hijacking and some legitimate routing changes those methods suffer from many false positives.

Other proposals [23], [40], [50], [49] leverage active probing of networks together with passive monitoring to improve the detection by assessing the impact of BGP routing changes on the data plane. In order to study BGP hijacking spammers we use the SpamTracer methodology we introduced in [47] for collecting a comprehensive set of routing-level features about spam networks. In [38] Roughan et al. advocates that neither BGP nor traceroute measurements were designed to infer the AS-level connectivity of the Internet and capture the complex inter-AS relationships, hence all results inferred from such data can only be as accurate as the data. However, we try to balance this limitation by setting up our own data collection process allowing us to collect the most appropriate data for studying the routing-level behavior of spammers.

In 2006, Ramachandran et al. [35] introduced a new phenomenon called **"BGP spectrum agility"**, where they claimed to have observed, over a period of a few months, spam from a set of large (*i.e.,* /8) previously unannounced IP address blocks hijacked for a very short period of time (*i.e.,* less than one day). Later, Hu et al. [23] and Duan et al. [20] confirmed these observations. However, we have recently shown in [46] through a practical case study that correlating BGP abnormal events with malicious network traffic is insufficient to conclusively identify malicious BGP hijacks. Meanwhile, Schlamp et al. [39] described a unique case where a couple of IP address blocks were hijacked for months to perform malicious activities, such as spamming.

These few publications show the existence and the reality of BGP hijacks in the wild but the scarcity of the attacks observed since 2006 give the impression that this threat remains highly anecdotal and that no infrastructure seems to have been put in place by hackers to automatize efficiently, systematically, the launching of BGP hijacks. The results we present here after will portray a very different situation in which, every day, several BGP hijacks are taking place. Furthermore, we show that this is by no means a new phenomenon. Our data highlight that this has taken place for 18 months, without anyone noticing it apparently.

III. Experimental setup

We have set up a comprehensive experimental environment to study the BGP hijacking spammers phenomenon. The complete setup is depicted in Figure 1. Our goal here is to Ⓐ collect routing data related to spam networks, Ⓑ extract from this data IP address blocks exhibiting an abnormal routing behavior and retain the ones most likely indicating they might result from a BGP hijack, Ⓒ manually (in)validate each candidate hijack by taking advantage of external data sources, and finally Ⓓ investigate the root cause behind some validated malicious BGP hijacks to obtain new insights into hijacking spammers behavior.

Our experimental setup builds upon SpamTracer [47], a system designed for the collection and analysis of routing data related to spam networks. The assumption behind this

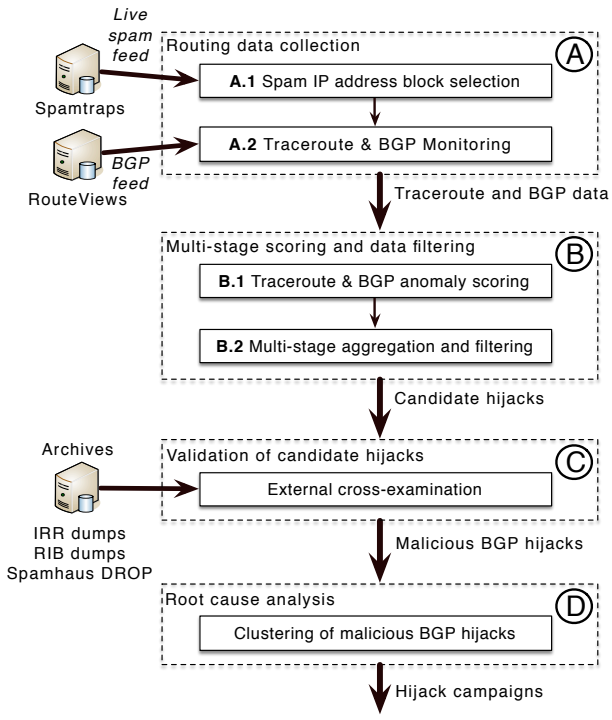

Fig. 1. Experimental environment.

approach is that when an IP address block is hijacked for stealthy spamming then a routing change will be observed when the block is released by the spammer to remain stealthy. Since we start monitoring a network when we receive spam from it, we look for a routing change from the hijacked state of the network to the normal state of the network. The goal here is not to build a stand-alone BGP hijack detection system but instead to collect, in real-time, routing data associated with spam networks in order to identify spammers sending spam from temporarily (*i.e.,* less than one day) hijacked IP address blocks as described in [35]. In the remainder of this section we describe the different parts of our experimental environment in more details.

A. Routing data collection

1) Spam IP address block selection: Our primary dataset is a live feed of spam emails collected at spamtraps. Every day we receive about 3,500,000 spam emails from about 24,000 distinct IP address blocks. Due to the overhead imposed by traceroute measurements and by querying the BGP collectors, our system can currently monitor about 8,000 IP address blocks on a daily basis. A sample of IP address blocks is extracted from the spam feed every hour. When selecting blocks to monitor we prioritize the *recently announced* ones as they are good candidates for short-lived hijacks as suggested in [35]. We consider to be *recently announced* any IP address block in our spam dataset that became routed within the last 24 hours, based on archived routing information bases (RIBs) from RouteViews and RIPE RIS.

2) Traceroute and BGP monitoring: We monitor each spam IP address block during seven days by launching traceroutes and collecting BGP AS paths towards the spam network from six RouteViews BGP collectors distributed worldwide. Additionally, we map every IP-level hop in traceroutes to the AS announcing it and further enrich the path with geolocation information [3] and the registered holder [11] of each IP- and AS-level hop traversed by a traceroute.

To summarize, every spam IP address block monitored over 7 consecutive days is associated with:

- a set of daily IP/AS traceroute paths from our vantage point to the network;
- a set of daily BGP AS paths from the six Routeviews BGP collectors to the network;
- additional information on the geolocation and owner of each intermediate IP- and AS-level hop in traceroutes.

B. Multi-stage scoring and data filtering

As the amount of IP address blocks to monitor increased significantly over time, we needed a mechanism to automatically investigate them. This is why we designed a multi-stage scoring and filtering system that analyzes the raw data, identifies abnormal routing events, assigns individual scores based on a consistent set of criteria, and then aggregates all scores to eventually highlight IP blocks most likely indicating possible BGP hijacks. We describe here the main components of this multi-stage scoring and filtering system.

1) Traceroute and BGP anomaly scoring: In [47] we introduced a set of heuristics that proved to be effective at identifying BGP and Traceroute anomalies. With the limited data we had at our disposal at that time (April-Sep 2011), this approach had only unveiled a limited number of cases involving a few IP blocks being hijacked. In this paper, we apply it to a much larger dataset and reveal a significantly larger amount of successful hijacks.

(I) BGP anomalies provide a view from the control plane on the routing behavior of monitored networks and are extracted from the set of daily BGP AS paths. (I.a) A *BGP origin anomaly* refers to an IP address block being announced by more than one AS. Such anomaly is also commonly referred to as a Multiple Origin AS (MOAS) conflict. (I.b) A *BGP AS path deviation* measures the difference observed between BGP AS paths collected from a given BGP collector towards a given IP address block.

(II) Traceroute anomalies are used to assess the impact of control plane routing changes on the data plane. They are extracted from the set of daily IP/AS traceroute paths. (II.a) An *IP/AS reachability anomaly* refers to a permanent change in the reachability of the probed destination host or AS. (II.b) A *Hop count anomaly* quantifies a possible permanent change in the length of traceroutes. (II.c) An *IP-level traceroute deviation* and an *AS-level traceroute deviation* measure the difference observed between respectively IP-level traceroutes and AS-level traceroutes. (II.d) A *Geographical deviation* quantifies the difference observed between the countries traversed by traceroutes.

As described in [47], every anomaly type is quantified with a score in $[0, 1]$. A BGP origin anomaly is defined by a

triplet (IP, AS_1, AS_2) where IP is the monitored IP address block and AS_1 and AS_2 are the ASes announcing IP. In case an IP address block is announced by more than two ASes, several BGP origin anomalies can be produced. Path deviations are computed using the Jaccard index[1] on the sets (p_d, p_{d+1}) where p_d is a path collected on day d and p_{d+1} is a path collected on day $d+1$. Finally, IP/AS reachability anomalies and the hop count anomaly are computed once for all traceroutes collected for a spam network. In summary, a network monitored for n days produces (i) zero or more BGP origin anomalies, (ii) $c \times (n-1)$ path deviations for each anomaly type where c is the number of collectors ($c = 1$ for traceroutes and $c = 6$ for BGP AS paths) and (iii) zero or one IP/AS reachability and hop count anomalies.

2) Multi-stage aggregation and filtering: In [47] we used a decision tree to compute a global suspiciousness score on the monitored IP blocks, based on various predefined combinations of routing anomalies. Thanks to the analytical experience we gained by investigating a large number of candidate hijack cases, we have further enhanced our anomaly scoring and filtering method by replacing the ad-hoc decision tree with a multi-stage aggregation system relying on Multi-Criteria Decision Analysis (MCDA) techniques. This new MCDA-based approach turns out to be more flexible, as well as easier to implement and maintain than a large set of ad-hoc rules. It is also more effective at assigning a global *suspiciousness score* to any given IP address block monitored by SPAMTRACER because it removes the need to define intermediate decision thresholds and allows to identify suspicious routing behaviors likely resulting from a BGP hijack in a more fine-grained fashion.

MCDA provides an extensive set of methods to model simple to very complex decision schemes, ranging from basic averaging functions to more advanced methods such as fuzzy integrals [17]. In our decision-making system, we rely mainly on the Weighted Ordered Weighted Average (WOWA) operator [45] to aggregate the different individual anomaly scores at various levels. The choice of using WOWA was motivated by a trade-off between flexibility and complexity of the decision model. In fact, WOWA combines the advantages of two types of averaging functions: the weighted mean (WM) and the ordered weighted average (OWA). This enables a decision maker to quantify, with a single operator, the reliability of the information sources (as WM does) but also to weight the individual scores according to their relative *ordering*. This sorting and weighted ordering aspects allow us to emphasize various distributions of scores (*e.g.,* eliminate outliers, emphasize mid-range values, ensure that "at least x" or "most of" the scores are significantly high, etc).

Obviously, like any other unsupervised technique (*i.e.,* in absence of reliable "ground truth" data), a number of parameters must be defined – usually based on the acquired expertise and domain knowledge – to accurately model a decision scheme and ensure that the most relevant cases are ranked in the top tier, whereas truly benign cases are assigned very low scores. In the case of WOWA, we only have to specify two different weighting vectors, which already simplifies considerably the parameter selection phase. This said, it is important to stress that the primary goal of our multi-stage scoring and filtering approach is to narrow down, as much as possible, the number of cases and be able to focus on a limited set of most promising BGP hijack candidates, which can be further validated through manual investigation. Recall that the ultimate goal is to prove whether (i) "BGP spectrum agility" still exists and (ii) the modus of BGP hijacking spammers has changed since 2006 [35]. In other words, we try to understand if this is a problem still worth of consideration in 2014, or not. Under these considerations, and without discrediting the importance of parameters selection, we argue that the determination of the *optimal* parameters for our decision model is, at this stage, not critical to achieving our goals.

We refer the interested reader to the Appendix to learn more details on the mathematical background behind our MCDA scoring and aggregation system as well as its parameters.

C. Validation of candidate hijacks

Due to the lack of ground truth information and the limitations of routing data alone to identify instances of BGP hijacks, an additional validation is required and consists in collecting additional evidence, usually involving some manual processing, about candidate hijacks to help confirm them or not. We (in)validate candidate hijacks using, besides the collected routing data, daily archives of the following external data sources:

- **Routing Information Base (RIB)** dumps from RIPE RIS [8] and Routeviews [14] consist of snapshots of routers routing table providing the list of announced IP address blocks and associated BGP AS paths.
- **Internet Routing Registry (IRR)** dumps [5] provide registration information on IP address and AS number holders as well as possible routing policies established between interconnected networks (*i.e.,* via BGP `import` and `export` rules).
- **Spamhaus Don't Route Or Peer (DROP)** [10] is a blacklist of IP address blocks allegedly controlled by cybercriminals, including some claimed to have been stolen from their legitimate owner.
- **Network operational mailing lists**, such as [7], [9], are sometimes used by network operators to report BGP hijack incidents (*e.g.,* the Link Telecom hijack [16]).

We examine the routing history related to candidate hijacked IP address ranges to study their routing characteristics including (i) when they were publicly announced, (ii) the BGP origin ASes used to advertise them, and (iii) the upstream provider ASes seen in the AS paths. Because our data collection system only collects routing information about IP address ranges for a limited period of time, we built the routing history of candidate hijacked IP address ranges from the archived dumps of routing information bases (RIBs).

We leverage IRR dumps to identify the country of registration and, the name and the contact details of the owner of IP address blocks and AS numbers involved in candidate

[1]The Jaccard index J of two sets S_1 and S_2 measures the amount of overlap between the two sets and is defined as $J = \frac{|S_1 \cap S_2|}{|S_1 \cup S_2|}$.

hijacks. We use this information to assess the legitimacy of routing announcements and profile IP address block and AS number holders, *e.g.,* to determine whether the owner of an IP address block is also the owner of the originating AS or to determine whether the owner of an announced IP address block is still in business. As suggested in [41], we further assess the consistency of inter-AS links observed in BGP AS paths using the published routing policies when available. We consider an inter-AS link consistent if both AS refer to each other in their declared `import`/`export` rules.

We use feedback from the Spamhaus DROP list that is a subset of SBL consisting of "IP address blocks that are hijacked or leased by professional spam or cybercriminal operations" [10].

Finally, in order to facilitate the communication among network operators in the Internet, the operational community uses public mailing lists, such as the North American Network Operators' Group (NANOG) mailing list [7] or the RIPE Working Groups mailing lists [9]. We check our candidate hijack cases against reported routing incidents in the archives of these two mailing lists.

At the end of this stage, we should be left with a set of hijack cases that should allow us to confirm or not[2] that BGP hijacks as described in [35] are still ongoing and, if yes, what their characteristics are.

D. Root cause analysis

While the external cross-validation of candidate hijacks described here above should increase our confidence in the existence of BGP spectrum agility spammers in the real world, we wanted to confirm our results by further investigating the root causes of the validated hijacks from a spam campaign perspective. Assuming we could identify good candidate hijacks that are perfectly matching the anomalous routing behavior of BGP spectrum agility spammers, one would expect that spam campaigns launched from these hijacked networks, by the same group of agile spammers, should intuitively share also a number of commonalities with respect to spam features (*e.g.,* advertised URI's, sender's address, etc).

We have thus used a multi-criteria clustering framework called TRIAGE [42] to identify series of spam emails sent from different hijacked IP address blocks that seem to be part of a campaign orchestrated by the same agile spammers. TRIAGE is a software framework for security data mining that relies on intelligent data fusion algorithms to reliably group events or entities likely linked to the same root cause. Thanks to a multi-criteria clustering approach, it can identify complex patterns and varying relationships among groups of events within a dataset. TRIAGE is best described as a security tool designed for intelligence extraction and attack investigation, helping analysts to determine the patterns and behaviors of the intruders and typically used to highlight how they operate. This novel clustering approach has demonstrated its utility in the context of other security investigations, *e.g.,* rogue AV campaigns [18], spam botnets [44] and targeted attacks [43].

[2]Disclaimer: We acknowledge that some inadequacies in the data exist leading to false positives. Nevertheless, as shown later, the pattern coming out of our dataset builds to a very compelling case.

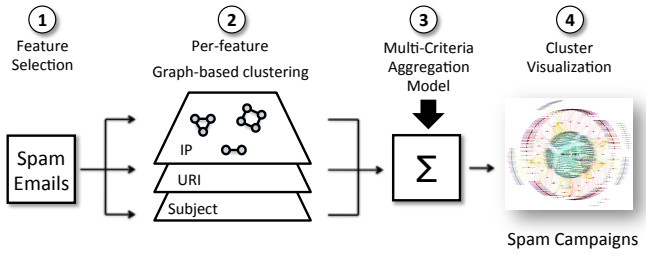

Fig. 2. Clustering spam emails sent from hijacked networks using TRIAGE.

Figure 2 illustrates the TRIAGE workflow, as applied to our spam dataset. In step ①, a number of email characteristics (or *features*) are selected and defined as decision criteria for linking related spam emails. Such characteristics include the sender IP address, the email subject, the sending date, the advertised URL's and associated domains and `whois` registration information. In step ②, TRIAGE builds relationships among all email samples with respect to selected features using appropriate similarity metrics. For text-based features (*e.g.,* subject, email addresses), we used string-oriented similarity measures commonly-used in information retrieval, such as the Levenshtein similarity and *N-gram* similarity [27]. However, other similarity metrics may be defined to match the feature type and be consistent to analyst expectations (*e.g.,* Jaccard to measure similarity between sets, or a custom IP addresses similarity metric that is based on their relative inter-distance in the binary space).

At step ③, the individual feature similarities are fused using an aggregation model reflecting a high-level behavior defined by the analyst, who can impose, *e.g.,* that some portion of highly similar email features (out of n available) must be satisfied to assign different samples to the same campaign (regardless of which ones). Similarly to the WOWA aggregation method explained here above, in TRIAGE we can assign different *weights* to individual features, so as to give higher or lower importance to certain features. For this analysis we gave more importance to the *source IP addresses*, *domain names* associated to spam URL's and `whois` registration names, since we anticipate that a combination of these features convey a sense and possible evidence of colluding spam activities.

As outcome (step ④), TRIAGE identifies *multi-dimensional clusters* (called MDC's), which in this analysis are clusters of spam emails in which any pair of emails is linked by a number of common traits, yet not necessarily always the same. As explained in [42], a decision threshold can be chosen such that undesired linkage between attacks are eliminated, *i.e.,* to drop any irrelevant connection that could result from a combination of small values or an insufficient number of correlated features.

IV. RESULTS

We now turn to the description of our results, by detailing step-by-step the outcome of every component of our experimental environment in Figure 1. We finish with a thorough investigation and validation of the candidate malicious BGP hijacks we have identified.

Statistic	Jan 2013-Jun 2014
Nr of distinct IP address blocks	391,444
Nr of distinct ASes	18,252
Nr of traceroutes	5,594,164
Nr of BGP viewpoints	6
Nr of BGP AS paths	25,679,725

TABLE I. SUMMARY OF THE BGP AND TRACEROUTE DATASET.

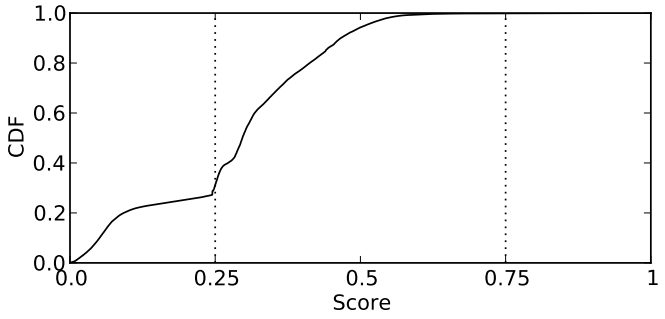

Fig. 3. BGP hijack identification: scores between January 2013 and June 2014.

A. Data collection results

We consider a dataset of BGP and traceroute data collected between January 2013 and June 2014 (1.5 years). A summary of the dataset is provided in Table I.

During 18 months we monitored a total of 391,444 distinct IP address blocks which sent spam to our spamtraps. These networks were operated from 18,252 different ASes. Finally, more than 5.5M data plane measurements and about 25.6M BGP routes towards these networks were collected.

B. Multi-stage scoring and filtering results

Figure 3 shows the distribution of scores for the monitored spam networks. The first part of the curve between the score value 0 and approximately 0.25 corresponds to 31.29% of spam networks exhibiting almost no variability in collected BGP routes and traceroutes. These are very likely benign cases. 68.60% of spam networks have a score between 0.25 and 0.75. Networks in that category usually exhibit a set of various anomalies, which makes them hard to attribute to a benign or malicious routing behavior. They may suffer from limitations of the aggregation model or from inaccuracies in the collected data [32], which, in the case of a benign routing behavior, mistakenly increases the suspiciousness score and, in the case of a malicious routing behavior, prevents it from being correctly extracted by our scoring system. Finally, 0.11% of monitored spam IP address blocks have a score higher than 0.75 and correspond to the most likely hijacked networks.

There are 437 different IP address blocks which exhibit a score higher than 0.75. Each of them was monitored only once during the 18 months of the experiment.

C. Validation of candidate hijack results

We leverage here the methodology presented in Section III-C to (in)validate uncovered candidate hijacks. Due to the large amount of time required to manually investigate cases, we focused our analysis on the 437 spam IP address blocks that scored above 0.75 in our multi-stage scoring and filtering system, *i.e.,* the upper quartile in the scoring distribution. Manual in-depth analysis of these 437 cases reveals that 373 of them are benign cases for one of the following reasons. First, most benign cases correspond to networks which disappear from the routing tables for a few hours to several days while the network is monitored and reappear after the network stopped being monitored thus creating different anomalies. This result is due to a limitation of our system, which monitors networks for a limited time period of seven days. Second, other cases were mistakenly flagged as suspicious due to inaccuracies in traceroute measurements, such as traceroutes cluttered with many non-responsive IP hops ("*"). Consequently, we concluded that 64 cases out of 437 were found to have likely been hijacked[3] for the reasons detailed below.

We found out that all these 64 remaining cases shared a common routing behavior: they appeared to be routed during the first one to six days of the monitoring period and then permanently disappeared from the routing tables. As a result all these cases exhibit similar routing anomalies triggered by the significant difference in control plane and data plane routes at the time and after the block was routed.

64 IP address ranges were found to have been hijacked between January 2013 and June 2014. After examining the routing history of these blocks, we could classify them further into two different categories:

- PREFIX HIJACK VIA VALID UPSTREAM: In 92% of the hijacks, the IP address ranges were allocated but ***(1) unannounced*** by the time they were hijacked (*i.e.,* left idle by their valid owner), and the attacker forged part of the BGP AS path to advertise the IP ranges using an ***(2) invalid BGP origin AS*** via a ***(3) valid direct upstream provider (first hop) AS***.
- AS HIJACK VIA ROGUE UPSTREAM: In 8% of the hijacks, the IP address ranges were allocated but ***(1) unannounced*** and the attacker forged part of the BGP AS path to advertise the IP address ranges using the ***(4) valid BGP origin AS*** but via an ***(5) invalid direct upstream provider (first hop) AS***.

(1) Unannounced IP address space: The routing history revealed that all hijacked prefixes were unannounced before being hijacked.
(2)-(4) (In)valid BGP origin AS: In this work, we consider the origin AS for an IP address range as *valid* if the IP address range is mapped to the origin AS in the IRR's (`whois`) and the IP address range owner is also the same as the origin AS owner.
(3)-(5) (In)valid direct upstream provider AS: In this work, we consider as *invalid* the AS a_1 appearing as the direct upstream provider of the origin AS a_0 in the BGP AS path $\{a_n, ..., a_1, a_0\}$ if all the following conditions are met: (1) it has never been used as a direct upstream provider AS for

[3]Disclaimer: In the remainder of the paper, for the sake of conciseness, we talk about hijacks and attacker instead of candidate hijacks and likely attacker even though we have no bullet proof evidence of their wrong doing.

a_0 in the past, (2) it does not appear in the list of provider ASes of a_0 and does not have a_0 in the list of its customers (*i.e.,* `imports`/`exports`) published in the `whois` when such information was provided, (3) it is not used as an upstream provider to advertise any non hijacked IP address range at the time it is observed in the hijacks, (4) it is unused when it is observed for the first time in hijacks, (5) its holder refers to an inactive organisation, and (6) it has been reported as suspicious by Spamhaus[4].

In the AS hijack cases, it thus appears that attackers actually forged part of the BGP AS path ($\{a_1, a_0\}$) by unauthorisedly using a_1 and a_0 in the BGP announcements for the different hijacked IP prefixes. BGP hijacking using a forged AS path is a stealthy BGP hijack technique [23], [39] and was probably used by the attackers in an effort not to raise suspicion.

We also mined archives of the NANOG [7] and RIPE Working Groups [9] mailing lists for public reports related to ASes or IP address blocks identified in our hijacks. We found only one thread [4] reporting the hijack of the block 91.220.85.0/24 and its legitimate BGP origin AS51888 via the invalid direct upstream provider AS42989.

We further observed that the 64 hijacked IP address blocks were advertised from only seven invalid distinct BGP origin ASes (prefix hijacks) and via only three invalid distinct upstream ASes (AS hijacks). Based on this observation we used the archived routing information bases (RIBs) from RouteViews and RIPE RIS to extract all IP address ranges originated by the same seven invalid BGP origin ASes or advertised via the same three invalid upstream provider ASes during the same 18 month time period. Surprisingly no less than 2,591 additional IP ranges were uncovered, all of them matching the exact same hijack signature as the other 64 IP address blocks. While we observed spam coming from the 64 hijacked IP address ranges identified by our system, we did not find any spam sent from the new 2,591 ranges in our spamtrap logs. In the remainder of this section we thus investigate **a total of 2,655 IP address ranges supposedly hijacked between January 2013 and June 2014**.

Figure 4 shows the distribution of the 2,655 observed hijacks across time. We can see that 95.3% of the observed hijacks have occurred after July 2013. From that point the distribution becomes almost uniform, showing that hijacks were performed on a regular basis for more than one year. With an average of 4.82 hijacks per day, we note that BGP hijacks have been an ongoing and recurring threat in the past 18 months (and possibly before or after).

We now focus on another key characteristic of the identified hijacks: their duration. In [35] Ramachandran et al. report on spam coming from IP prefixes involved in BGP routing announcements lasting less than one day. In our case, 85.5% lasted **less** than one day, from 29 minutes to 23 hours 47 minutes, 94.6% lasted no more than 2 days and 98.7% lasted no more than one week. A large fraction of hijacks are thus

[4]Spamhaus SBL records related to some identified hijacked IP address blocks are available at http://www.spamhaus.org/sbl/query/SBL<record_id>: 96354, 175835, 177177, 177452, 177570, 179312, 180606, 182044, 182223, 182351, 183715, 183836, 184596, 184865, 185726, 185728, 217199. Note that records are purged when the cases are considered to be solved.

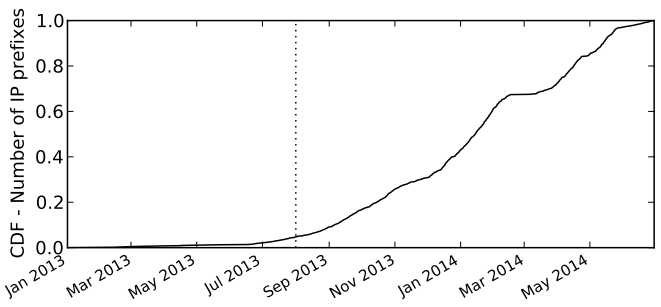

Fig. 4. The number of hijacked IP address ranges observed between January 2013 and June 2014. Most of the observed hijacks occurred after July 2013.

similar in duration to those reported in [35], *i.e.,* less than one day. All in all the majority of hijacks are rather **short-lived**. Additionally, 1.3% of hijacks are **long-lived**, *i.e.,* lasted more than one week in our observations, with a maximum duration of 8 months and 20 days.

Though short-lived and long-lived hijacks share some characteristics, we consider them to be due to two distinct phenomena. In fact, short-lived hijacks can be used by an attacker to circumvent traceback and avoid blacklisting by hopping between IP addresses in a range until the range itself gets blacklisted and then move to another range. Long-lived hijacks however make it harder for the attacker to remain undetected. We prove this later when checking the list of hijacked IP address ranges against several blacklists. Such long-lived hijacks have already been observed in the wild, for instance in 2011 a couple of IP prefixes belonging to the company Link Telecom were hijacked for 5 months and used to perform various malicious activities such as sending spam and hosting services but also exploiting remote hosts and originating suspicious IRC traffic [39].

In the remainder of this Section we investigate short-lived ($\leq$ 1 week) and long-lived ($>$ 1 week) hijacks separately to emphasize their similarities and differences. We consider the following characteristics of a hijack event:

- (*C.1*) Whether **spam emails** were received from the IP address range at our spamtraps and/or spam sources were **blacklisted** for the IP address range in Spamhaus SBL or DROP (Don't Route Or Peer) [10], Uceprotect [13] or Manitu [6].
- (*C.2*) The **duration of the unadvertised period** of the IP prefix, which corresponds to the amount of time elapsed between the last time it was announced and the moment it was hijacked.
- (*C.3*) The **registration date** of the IP address range, which is the date at which it was allocated or assigned by a Regional Internet Registry (RIR) to an Internet Service Provider (ISP) or end-user (*e.g.,* a company).
- (*C.4*) The **size** of the IP address range, which defines the number of individual IP addresses available in the range.
- (*C.5*) Whether the **owner** of the IP address block is still in business.

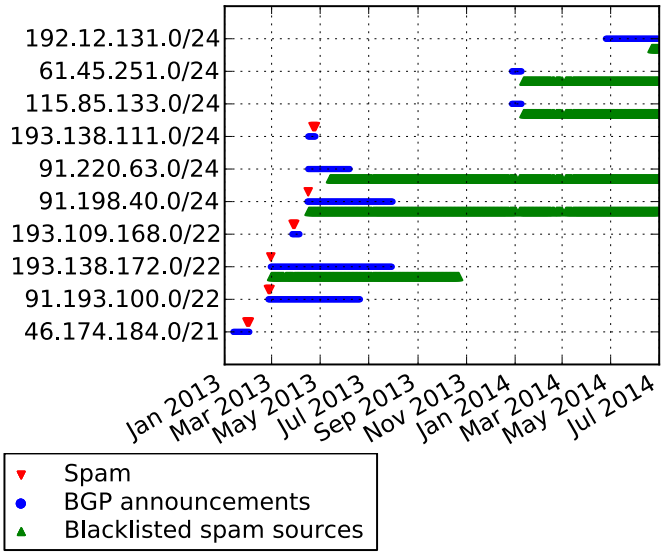

Fig. 5. BGP announcements, spam emails and blacklisted spam sources related to **long-lived** hijacked IP address ranges. For the sake of conciseness, only the 10 out of 35 IP address ranges that sent spam to our spamtraps or were blacklisted are depicted.

1) Long-lived hijacks

In this section we analyse more closely the 35 long-lived hijacks (out of the total 2,655 hijacks) we identified according to the *five* characteristics described above.

(C.1) Figure 5[(5)] shows the spam and blacklisted spam sources along with BGP announcements related to long-lived hijacked IP prefixes. Since those IP prefixes were not announced before or after being hijacked, the BGP announcements shown here all relate to the time of the hijacks. We can see that six out of 35 IP address ranges sent spam to our spamtraps. A total of 815 spam emails were sent from IP addresses scattered throughout each of the long-lived hijacked IP address blocks. Spam was mainly received at the start of the hijack period. No IP source was found to be blacklisted at the time the spam was received.

However, two networks (193.138.172.0/22 and 91.198.40.0/24) out of the 35 became blacklisted by Spamhaus within two days after they became hijacked and we observed spam originating from them. In these cases, blacklists appear to have reacted quickly. Four additional networks which have not sent spam to our spamtraps also became blacklisted, although it took more time for them to appear on a blacklist. For two of them (61.45.251.0/24 and 115.85.133.0/24) it took 2 weeks and the hijack was over by the time they appeared on a blacklist. For the other two (91.220.63.0/24 and 192.12.131.0/24) it took one month and 2 months respectively before they appeared on a blacklist.

(C.2) 26 IP prefixes out of 35 were never announced on the Internet before they were hijacked. The 9 others were hijacked on average one year after remaining unadvertised for at least one day, and maximum three years and two months.

[5]Disclaimer: IP address blocks and ASes were likely abused in hijacks between January 2013 and June 2014 and, therefore, might now be legitimately used.

(C.3) The 35 long-lived hijacked IP address blocks were mostly registered after 2000. It is noteworthy that at the time they were hijacked, these ranges were all properly registered IP address blocks assigned to an organisation. None of them was part of "bogon" IP address blocks, *i.e.,* IP addresses that should not be announced on the Internet [11].

(C.4) In [35], Ramachandran et al. claimed to have observed spam from large (*i.e.,* /8) hijacked IP address blocks. In our 35 long-lived hijack cases, the IP address blocks were smaller than what was claimed in previous studies, *i.e.,* the largest was a /19 and the smallest was a /24.

(C.5) The analysis of `whois` records for long-lived hijacked IP address blocks revealed that most of the 35 blocks refer to organisations that are apparently out of business. This observation indicates that attackers might specifically target unannounced IP address space whose registrant does not exist anymore, for instance when a company is dissolved, acquired by or merged into another one. In some cases, its IP address blocks may be left unused.

Last but not least, we managed to get feedback from an ISP unwittingly involved in 23 out of the 35 long-lived hijacks. After investigation on their side, the ISP confirmed these attacks had taken place and were performed by one of their customers, without them noticing it initially. The elements we provided corroborated their observations, and the ISP has since then terminated his peering contract with the misbehaving AS owner.

2) Short-lived hijacks

In this Section we focus our analysis on the 2,620 short-lived hijacks (out of the total 2,655 hijacks). We further distinguish two episodes in these short-lived hijacks: (1) spam and blacklisted spam sources related to hijacked networks observed between February and May 2013 and (2) an interesting hijack phenomenon observed between June 2013 and June 2014, showing a striking and unusual temporal pattern in the BGP announcements. We first present these two episodes and their differences with respect to characteristic C.1. Afterwards, we describe their commonalities in terms of the other characteristics C.2-5.

Episode 1: From February until May 2013

(C.1) Out of 2,620 short-lived hijacked IP prefixes, 58 have sent spam emails to our spamtraps between February and May 2013. Figure 6 shows the BGP announcements, spam and blacklisted spam sources related to a sample of 25 out of 58 short-lived hijacked IP prefixes. The figure highlights:

- the strong **temporal** correlation between BGP announcements and spam, and
- the **low** number of IP address blocks (7 out of 58) blacklisted by Spamhaus before the end of the hijack.

A total of 4,149 spam emails were received from the short-lived hijacked IP address blocks. We extracted from this spam all advertised URLs that were pointing to 1,174 unique

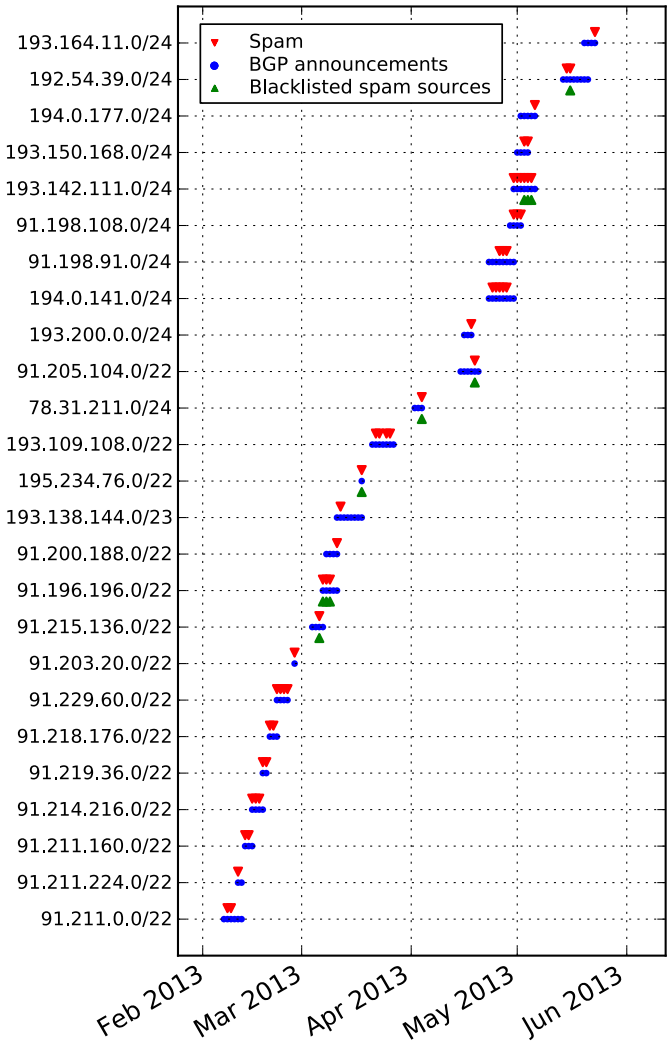

Fig. 6. Episode 1 of **short-lived** hijacks between February and May 2013: temporal correlation of BGP announcements, spam emails and blacklisted spam sources related to hijacked IP address ranges. For the sake of conciseness, only a sample of 25 out of 58 IP address ranges are depicted.

domain names, resolving to IP addresses belonging to the same hijacked IP address blocks, showing that some IP addresses were used in parallel to send spam and host the advertised scam websites. From `whois` information, we observed that these domain names were usually created within a few days before the networks being hijacked. This shows that attackers, very likely, control the entire IP address blocks and take full advantage of them.

Furthermore, spam emails collected by our spamtraps are enriched with the name of the spambot associated with the spam, by taking advantage of CBL signatures [2]. Spam emails sent from the supposedly hijacked IP address blocks we uncovered were not associated with any known spam botnet and thus must have been sent using another type of spamming infrastructure, instead of the traditional spam botnets. This is consistent with BGP spectrum agility where spammers need to set up a dedicated infrastructure with their own machines so that they can be assigned the hijacked IP addresses.

Episode 2: From June 2013 until June 2014

While examining the short-lived hijacks in the first period, we uncovered an intriguing phenomenon in the hijacks performed during the second period, between June 2013 and June 2014. This phenomenon is significant since it includes 2,562 short-lived hijacks representing 97.8% of all short-lived hijacks identified. Figure 7 depicts a sample of 87 (out of 2,562) hijacks that occurred in June 2014 and shows that:

- all hijacks are actually performed by groups of two to four prefixes, starting and ending at the same time;
- during the one month period there are always, at any point in time, at least two IP prefixes hijacked.

Although only part of the phenomenon is depicted in Figure 7, it is recurrent and persistant over the complete 13 month period, between June 2013 and June 2014[6]. This strongly indicates that they may have been performed with the same modus operandi. The fact that some groups of hijacks start only seconds after the end of previous groups further suggests that they might be carried out in an **automated way**, possibly also relying on some automated process to find target network address blocks to hijack.

(C.1) Strangely enough, we have not been able to find any malicious traffic associated with those hijacked IP address blocks. The absence of spam and other scam-related traffic in our data may be due to incomplete visibility into malicious activities associated with these networks, or could indicate that this is a moving infrastructure to host servers, *e.g.,* C&C servers. We have currently no conclusive evidence to validate this conjecture, though.

Common characteristics of episodes 1 and 2

In this section, we analyze common characteristics of all 2,620 short-lived hijacks.

(C.2) Figure 8 presents the duration of the unadvertised period of all short-lived hijacked networks. 2,261 IP prefixes (86.3%) were never announced before they were hijacked. From an informal discussion with a RIPE NCC executive [19] it is apparently common practice for network operators to register and use publicly routable IP address blocks for internal network infrastructure. This could explain why no route to such block can be found in our BGP feed. Apart from this reason we are not aware of any other reason why IP address blocks are registered but never actually announced. The remaining 359 networks were last announced from 6 days to 4 years before being hijacked with a average of 24.6 months and a median of 24.5 months. With 72.4% of IP prefixes left unannounced for more than one year we can conclude that attackers mostly hijack networks left unannounced for a long period of time.

(C.3) It appears that 1,775 IP address ranges (67.8%) out of 2,620 were registered before 1997 when the RIR's started taking on the registration of IP address resources and setting up the IRR's. Network address ranges registered before 1997

[6]The figure depicting the complete phenomenon of episode 2 is available at http://bit.ly/ndss2015_bgphijacks_episode2.

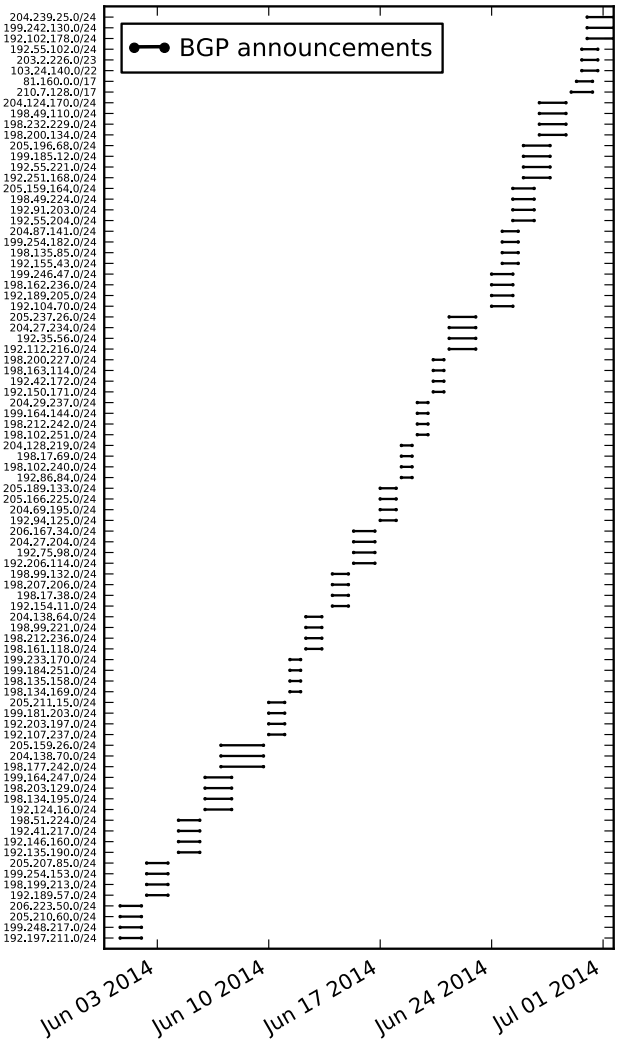

Fig. 7. Episode 2 of **short-lived** hijacks between June 2013 and June 2014: hijacks are always performed by groups of at least two IP prefixes. For the sake of conciseness, only a sample of 87 (out of 2,562) IP address ranges hijacked in June 2014 are depicted.

can thus sometimes be poorly documented and, for that reason, has been considered to be a target of choice for spammers to hijack them [22], [33]. This idea appears to be supported by our data.

(C.4) Short-lived hijacked IP address blocks include /17's, /21's, /22's, /23's and (92.6%) /24's, similar to the long-lived ones. Although those hijacks look like the ones Ramachandran et al. in [35] reported, the average size of hijacked address blocks is very different, namely /24, instead of /8.

(C.5) The analysis of `whois` records (from IRR databases) of short-lived hijacked networks revealed that all IP address blocks were, at the time they were hijacked, properly registered blocks assigned to an organisation with sometimes multiple blocks referring to the same organisation. Although we could not check all 2,620 IP address blocks, we looked at 100 of them and determined that 41% refer to organisations that are apparently out of business but, interestingly, 59% refer to organisations that appear to be still in business.

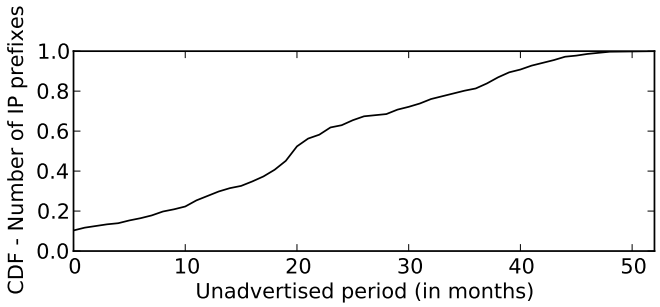

Fig. 8. The duration of the unadvertised period for 13.7% of the **short-lived** hijacked IP address ranges. The remaining 86.3% of IP address ranges were never announced before being hijacked.

Finally, 770 out of the 2,620 short-lived hijacks were later confirmed by an ISP, unwittingly involved in these attacks. After receiving a lot of complains for spam, the ISP discovered that one of their customers was indeed announcing IP address blocks he did not own, which were indeed used for nefarious activities as confirmed by our observations. The ISP has now terminated their contract with the misbehaving customer (who turned out to be colluding with a spammer apparently based in Russia).

D. Root cause analysis results

In the previous section we have uncovered strong evidence of BGP spectrum agility occurring in the Internet. However, we have not systematically analysed if the identified hijacks are isolated attacks or if some of them share a common root cause, as we would expect if they are part of campaigns orchestrated by the same spammers. This is why we have run the TRIAGE clustering tool against all spam emails coming from the 64 supposedly hijacked IP address blocks which have sent spam to our spamtraps.

The multi-criteria clustering tool has identified only 30 multi-dimensional clusters (MDC's) in which spam emails are correlated by various combinations of features. Because of the way these clusters are generated, we anticipate they likely represent different campaigns organized by the same individuals - as spam emails within the same cluster share several common traits. Thus 64 prefixes were used to run 30 different spam campaigns. In the following we will show that some campaigns are rather short-lived and run from a single prefix whereas others last for several days relying on a number of different prefixes. Table II provides global statistics computed across all MDC's. Most spam campaigns seem to be short-lived (lasting on average only a couple of days), except two MDCs that existed for more than 30 days.

By clustering spam emails into campaigns, we obtain new insights into hijacking spammers behavior. From the structure of MDCs, we uncover three key modus operandi of hijacking spammers: (1) 10 campaigns (out of 30) involve a single hijacked IP prefix that is not abused elsewhere in any other campaign, (2) 17 campaigns involve a single hijacked IP prefix, yet the hijacked prefix is abused concurrently in different spam campaigns, and (3) three campaigns were observed abusing *multiple hijacked IP prefixes* sequentially over a longer period

Statistic	Avg	Med	Min	Max
Nr of spam emails	141.8	11.5	2	1,178
Nr of IP prefixes	1.6	1	1	12
Nr of URL hosting server IP addresses	7.3	4	1	24
Nr of URL domain names	10.3	2	1	173
Nr of URL domain name whois registrants	44.5	6.5	1	556
Nr of spam subjects	47.7	7	2	455
Nr of active days	5.7	1	1	24
Lifetime in days	5.7	1	1	81
Compactness	0.43	0.43	0.27	0.74

TABLE II. GLOBAL STATISTICS FOR THE 30 MD-CLUSTERS (SPAM CAMPAIGNS).

of time. While the first two phenomena actually confirmed our intuition about the anticipated behavior of this class of spammers, the latter phenomenon is the most interesting as it confirms the existence of BGP spectrum agility in the form of campaigns of BGP hijacks orchestrated by the same spammers. Indeed, it highlights the existence of a more agile and sophisticated modus operandi of spammers capable of hijacking and abusing multiple IP prefixes, and subsequently hopping from one hijacked IP prefix to another to distribute spam. This agility enables them to send spam in a more stealthy manner and thus stay undetected "under the radar".

Finally, from Table II we also observe a higher variability in spam email subjects and `whois` registrant addresses, suggesting that spammers have automated tools at their disposal to facilitate the creation of new email templates and automate the registration of new domains used for disposable "one-time URL's".

Figure 9[7] shows a graph visualization of one of the large-scale campaigns that involved multiple hijacked IP prefixes, which illustrates the typical modus operandi of agile spammers operating such stealthy campaigns. In this particular example, we can observe the following key points:

- over 662 spam emails have been sent from 12 different hijacked IP prefixes (yellow nodes), each of them used in turn by spammers to distribute spam using a bunch of one-time URL's, most of them including domain names (blue nodes) registered at ENOM (large pink node) using privacy-protected email addresses provided by *whoisprivacyprotect.com* (red nodes);
- spam advertised content (domain URL's) share the same server IP addresses (lightgrey nodes);
- the campaign has a lifetime of 84 days, yet only 24 active days (purple nodes laid out in a clockwise fashion), during which spammers are hopping from one hijacked IP prefix to another, which is an effective way of circumventing IP-based spam filters and reputation systems.

To the best of our knowledge, these results are completely novel and shed a new light on the behavior of agile BGP hijacking spammers. First, we observe that stealthy spam campaigns can be performed by exploiting multiple hijacked IP address blocks. Secondly, we observe that the same invalid direct upstream providers were involved in all these hijacks, which already gives some indication of possible counter-measures. Finally, all URL's advertised in spam emails are sharing a common hosting infrastructure and were registered in a similar way – suggesting that `whois` registration data can also be leveraged in prevention systems. The key take-away of this root cause analysis is that it enables us to link together different hijacked prefixes showing they are used by the same spamming actors for a long period of time in a very stealthy way.

[7]Disclaimer: IP addresses, domain names and email addresses were found in campaigns launched from likely hijacked networks only between January 2013 and June 2014. These may have been abused and stolen from their legitimate owners and, therefore, may now be legitimately used.

E. Summary

Finding 1: We uncovered two types of hijack phenomena: long-lived and short-lived. Long-lived hijacks can last from a week to several months, whereas short-lived hijacks last from a few minutes to several days.

Finding 2: Attackers were found to stealthily hijack properly registered but unannounced IP address space by using two different hijacking techniques (as defined in Section IV-C on page 6): *prefix hijacking* and *AS hijacking*. In prefix hijacking, the attacker announced an IP address block using an invalid BGP origin AS via a valid direct upstream provider (first hop) AS. In AS hijacking, the attacker announced an IP address block using its valid BGP origin AS but via an invalid direct upstream provider (first hop) AS.

Finding 3: In the 2,454 *prefix hijacks* we found only *six different* invalid BGP origin ASes. In the 201 *AS hijacks* we found, for 195 different valid BGP origin ASes, only *three different* invalid upstream provider ASes. One AS, involved in the hijack of 793 IP address blocks over 16 months, was observed first as an invalid upstream provider AS, and then as an invalid BGP origin AS. These 793 hijacks were later confirmed by the ISP providing transit to that AS, who consequently terminated the contract with this customer abusing the routing infrastructure.

Finding 4: Spamming using hijacked IP prefixes appears to be an effective technique for defeating known protections, such as spam IP blacklists. Moreover, almost none of the IP address blocks were hijacked more than once meaning that in this case blacklisting those blocks after the hijack ends is not particularly useful. Finally, spammers also use the hijacked IP address blocks as a hosting infrastructure for spam advertised content.

Finding 5: Spammers mostly hijack IP prefixes that have never been advertised or left unadvertised for a very long time, typically more than one year.

Finding 6: Hijacking spammers seem to prefer IP address blocks that were properly registered, in contrast to "bogon" IP address blocks whose announcements are commonly automatically filtered out using for instance the list from Team Cymru [11].

Finding 7: Many hijacked IP address blocks we identified refer to organisations that ceased to exist. Orphan IP address blocks that are left behind then become targets of choice

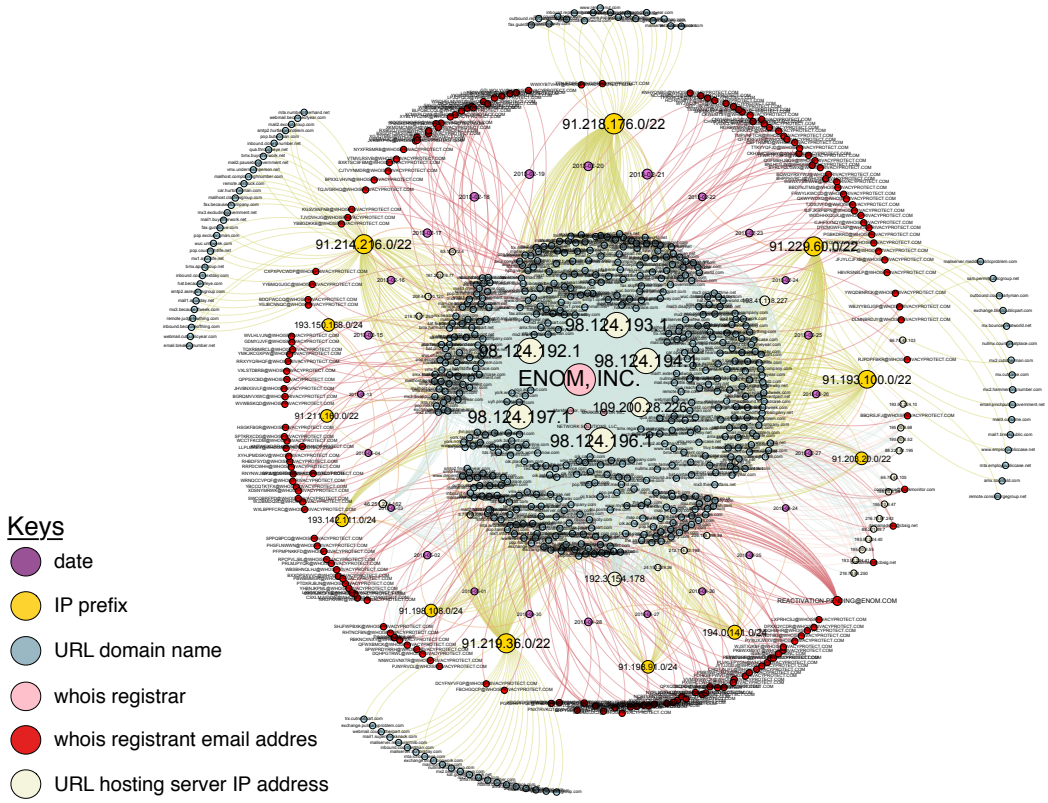

Fig. 9. An example of a large-scale spam campaign involving multiple hijacked IP prefixes. The nodes laid in clock-wise fashion reflect the timeline of the campaign.

for spammers as they can likely hijack them without being noticed. As of July 2014 as much as 20.26% of the whole IPv4 address space[8] is currently allocated or assigned but not publicly announced.

Finding 8: Some short-lived hijacks were clearly associated with spam activities, confirming the existence of the BGP hijacking spammers phenomenon as introduced in [35]. However, a large portion of them exhibited no spam and we conjecture that they would ideally serve as a moving infrastructure to host malicious servers.

V. Effectiveness of Current Counter-Measures

Different technologies and systems have been designed to detect and mitigate BGP hijacks. In this section we evaluate the effectiveness of two BGP hijack counter-measures: a state-of-the-art BGP hijack detection system called Argus [40] and the BGP security framework RPKI [24], [30], [31].

A. BGP hijack detection

There have been numerous systems [23], [29], [40], [1] developed to detect and mitigate BGP hijacks. One of them, Argus [40], aims at detecting BGP hijacks in real-time by using a combination of BGP data and ping measurements to detect, upon a routing change related to a network, changes in the network's reachability indicating a possible hijack. In an effort to assess the security impact of the hijack incidents we uncovered, we decided to verify the effectiveness of the Argus system against these cases. We chose Argus for two reasons: (i) it is currently deployed and publicly available, and provides an historical feed of alerts, and (ii) it is also able to detect all types of hijacks, namely those where the attacker hijacks an IP address block by using an invalid BGP origin AS or by forging part of the BGP AS path.

It turns out that none of the 2,655 hijacks we identified were reported by Argus. The reason is that most BGP hijack detection systems [23], [29], [40], [1] work by building a model of the Internet AS-level topology and then using it to validate any routing change. However, because all hijackings we identified involve IP space that was unannounced prior to being hijacked, there is no state for the IP address blocks in the model resulting in any new route announcement to be accepted as legitimate. Although current BGP hijack detection techniques are valuable for network operators to monitor their own networks, their inability to currently detect hijacks like those we observed suggest that those techniques should integrate in the future in their detection scheme some characteristics of the hijacks we have identified.

B. BGP hijack prevention

Besides BGP hijack detection techniques, the network operators have started to adopt and deploy a BGP hijack prevention framework commonly referred to as the RPKI system. Though many approaches have been proposed to

[8] Based on statistics published by RIR's and available at http://bgp.potaroo.net/ipv4-stats/prefixes.txt

bring security to BGP [25], this framework has been gaining more momentum than others in the last few years. This is likely due to the fact that it is the only framework being standardised by the IETF. We are not aware of any other framework that is ready and mature enough to go through that process. The framework relies on a Resource Public Key Infrastructure (RPKI), standardised in RFC 6480 [31], to prevent the injection of bogus routing announcements. The RPKI used in this scheme consists of a database of certificates of two types: (i) a type A called Route Origin Authorisation (ROA) binds an IP address block to its authorised BGP origin AS(es), and (ii) a type B that binds a router to the AS number it belongs to. The certification chain follows the AS number and IP address delegation chain, with the IANA acting as the root certificate authority for RIR's certificates, a RIR is then acting as the certificate authority for ISP's certificates, etc. Each certificate is signed with the private key of its holder and also embeds its public key. The framework proposes two separate techniques to secure BGP: (1) secured *route origination* and (2) secured *route propagation* (or BGPsec). Secured route origination, standardised in RFC 6483 [24], uses ROAs (type A certificates) to verify that a given IP address block is originated by the authorised AS(es). A router is then able to verify the validity of a received BGP update for a given IP address block and BGP origin AS by (1) querying the RPKI for a ROA related to the IP address block and verifying its cryptographic validity, and, (2) if the ROA is valid, verifying that the origin AS and the length of IP prefix observed in the BGP update match the authorised origin AS(es) and prefix length in the ROA. This prevents an attacker from announcing a block he does not own. Secured route propagation [30] aims at preventing *AS path forgery* by ensuring that each AS in the AS path was not impersonated. This is done by having each router signing a BGP update it propagates so that subsequent routers can verify, using type B certificates from the RPKI, that all routers which have signed the update indeed belong to the ASes found in the path.

Secure route origination is progressively being deployed. According to the RIPE NCC [37] there is currently 4.1% of the IPv4 address space covered by ROA's. Interestingly, none of the IP address blocks that we identified as having been hijacked were covered by a ROA at the time they were hijacked. In 92% of the hijacks we observed, the attacker announced the IP address blocks using an invalid BGP origin AS (*prefix hijacks* as defined in Section IV-C on page 6). Providing a ROA had been issued for these blocks and their valid BGP origin AS, the RPKI would have invalidated the bogus announcements.

However, assuming ROA's would bind IP address blocks with their legitimate BGP origin AS, hijacks can still be successful if attackers forge the BGP AS path and prepend the valid BGP origin AS to it, which is exactly what we observed in 8% of the hijacks (*AS hijacks* as defined in Section IV-C on page 6). Secured route propagation is currently still at an early stage and not yet being deployed. In the meanwhile, although BGP origin validation via ROAs does not intend to prevent BGP AS path forgery, as acknowledged in RFC 6483 [24], the RPKI and ROAs could nevertheless be leveraged to prevent unannounced IP address blocks from being hijacked by issuing a ROA for AS0 and each unannounced IP address block (such ROA's are already used to prevent the announcement of reserved/unallocated IP space as dictated in RFC 6483 [24]). Then, the RPKI will classify all routes for these IP address blocks as invalid. This solution is not perfect though as it requires a specific ROA to be issued when an IP address block becomes unannounced which, in the case of orphan blocks, is unlikely. Overall, the only proper solution to prevent BGP AS path forgery and the AS hijacks we identified is to have secured routed propagation, *i.e.,* BGPsec, deployed. Unfortunately, this solution is much more invasive and cannot be deployed without substantial software and hardware updates on all routers. Moreover, the standardisation process of BGPsec is not yet completed and there is no router code available as of today. Some vendors are working on it, or intending to work on it, but some other vendors do not even list it on their roadmap.

VI. Lessons learned and conclusion

We conclude by providing concrete lessons that can be leveraged to improve existing spam and BGP hijack mitigation techniques and thwart these attacks.

Lesson 1: We have confirmed the existence of BGP spectrum agility in the real-world in the form of stealthy and persistent campaigns of malicious BGP hijacks.

Lesson 2: Today's BGP hijack mitigation systems, such as [23], [29], [40], [24], [1], are **blind** to hijacks of registered though **unannounced** IP address space carried out by announcing an IP address block using its **valid** BGP origin AS but via an invalid upstream provider AS. The complete deployment of BGPsec and ROA's would prevent these attacks. In the meantime, we would suggest BGP hijack detection systems to include signatures for these hijacks based on the characteristics we uncovered.

Lesson 3: Owners of unannounced IP address blocks leave them vulnerable to hijacking. A best practice would be to announce all blocks even if they are unused.

Lesson 4: A worldwide hunt for orphan IP address blocks should be launched to prevent them from being hijacked and further used for malicious purposes. Additionally, IP address block owners that cease to exist or do not require the IP resources anymore should (be forced to) return them. Keeping IRR and RPKI data fresh is therefore key to prevent hijacks of such IP address space.

Lesson 5: Uncovered hijacks involved many different IP address blocks and origin ASes but very few invalid BGP origin ASes and direct upstream provider ASes. This suggests that ASes identified as invalid or malicious in previous hijacks can be leveraged to identify subsequent hijacks or even block traffic from and to IP address blocks advertised via these ASes.

As future work we plan to expand the collaboration we have recently initiated with CERT's, ISP's and the NANOG and RIPE communities at large. A concrete outcome of these ongoing discussions was the confirmation that one of the ASes found to be malicious by our system and responsible for the hijack of 793 IP prefixes has seen his peering contract terminated by its valid upstream ISP.

REFERENCES

[1] "BGPmon Network Solutions Inc." http://www.bgpmon.net/.

[2] "Composite Blocking List," http://cbl.abuseat.org/.

[3] "GeoIP: MaxMind," http://www.maxmind.com/.

[4] "Illegal usage of AS51888 (and PI 91.220.85.0/24) from AS42989 and AS57954 (in ukraine)," http://mailman.nanog.org/pipermail/nanog/2013-May/058230.html.

[5] "IRR.net," http://www.irr.net/.

[6] "Manitu.net DNSBL," http://www.dnsbl.manitu.net/.

[7] "North American Network Operators' Group (NANOG) mailing list," https://www.nanog.org/list/.

[8] "RIPE Routing Information Service (RIS)," http://www.ripe.net/data-tools/stats/ris/.

[9] "RIPE Working Group Mailing Lists," https://www.ripe.net/ripe/mail/wg-lists/.

[10] "Spamhaus," http://www.spamhaus.org/.

[11] "Team Cymru Community Services," http://www.team-cymru.org/.

[12] "The New Threat: Targeted Internet Traffic Misdirection," http://www.renesys.com/2013/11/mitm-internet-hijacking/.

[13] "Uceprotect," http://www.uceprotect.net/.

[14] "University of Oregon RouteViews Project," http://www.routeviews.org/.

[15] "YouTube IP Hijacking," http://www.nanog.org/mailinglist/mailarchives/old_archive/2008-02/msg00453.html.

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

[17] G. Beliakov, A. Pradera, and T. Calvo, *Aggregation Functions: A Guide for Practitioners*. New York: Springer, Berlin, 2007.

[18] M. Cova, C. Leita, O. Thonnard, A. D. Keromytis, and M. Dacier, "An analysis of rogue AV campaigns," in *RAID*. Springer-Verlag, 2010, pp. 442–463.

[19] A. de la Haye, "Chief Operations Officer at RIPE NCC," RIPE67, October 2013.

[20] Z. Duan, K. Gopalan, and X. Yuan, "An Empirical Study of Behavioral Characteristics of Spammers: Findings and Implications," *Computer Communications*, vol. 34, no. 14, pp. 1764–1776, Sep. 2011.

[21] R. Hiran, N. Carlsson, and P. Gill, "Characterizing large-scale routing anomalies: a case study of the china telecom incident," in *PAM*. Springer-Verlag, 2013, pp. 229–238.

[22] M. Hogewoning, "IP Hijacking: Secure Internet Routing," March 2012.

[23] X. Hu and Z. M. Mao, "Accurate Real-Time Identification of IP Prefix Hijacking," in *Security and Privacy*. IEEE, 2007, pp. 3–17.

[24] G. Huston and G. Michaelson, "Validation of Route Origination Using the Resource Certificate Public Key Infrastructure (PKI) and Route Origin Authorizations (ROAs)," February 2012, RFC 6483.

[25] G. Huston, M. Rossi, and G. Armitage, "Securing BGP: A Literature Survey," *Communications Surveys Tutorials, IEEE*, vol. 13, no. 2, pp. 199–222, 2011.

[26] V. Khare, Q. Ju, and B. Zhang, "Concurrent prefix hijacks: occurrence and impacts," in *IMC*. ACM, 2012, pp. 29–36.

[27] G. Kondrak, "N-gram similarity and distance," in *Conf. on String Processing and Information Retrieval*, 2005, pp. 115–126.

[28] C. Kruegel, D. Mutz, W. Robertson, and F. Valeur, "Topology-Based Detection of Anomalous BGP Messages," in *RAID*, 2003, pp. 17–35.

[29] M. Lad, D. Massey, D. Pei, Y. Wu, B. Zhang, and L. Zhang, "PHAS: A prefix hijack alert system," in *USENIX Security Symposium*, 2006.

[30] M. Lepinski, "BGPSEC Protocol Specification," February 2013, internet-Draft.

[31] M. Lepinski and S. Kent, "An Infrastructure to Support Secure Internet Routing," February 2012, RFC 6480.

[32] Z. M. Mao, J. Rexford, J. Wang, and R. H. Katz, "Towards an accurate AS-level traceroute tool," in *SIGCOMM*. ACM, 2003, pp. 365–378.

[33] L. Nobile and L. Vegoda, "Address Space and AS Hijacking," http://meetings.ripe.net/ripe-48/presentations/ripe48-eof-nobile-vegoda.pdf, May 2004.

[34] J. Qiu and L. Gao, "Detecting Bogus BGP Route Information: Going Beyond Prefix Hijacking," in *SecureComm*. IEEE, 2007, pp. 381–390.

[35] A. Ramachandran and N. Feamster, "Understanding the Network-Level Behavior of Spammers," in *SIGCOMM*. ACM, 2006, pp. 291–302.

[36] Y. Rekhter, T. Li, and S. Hares, "A Border Gateway Protocol 4," January 2006, RFC 4271.

[37] RIPE NCC, "RPKI ROA certification statistics," http://certification-stats.ripe.net/.

[38] 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," *IEEE Journal on Selected Areas in Communications*, vol. 29, no. 9, pp. 1810–1821, 2011.

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

[40] X. Shi, Y. Xiang, Z. Wang, X. Yin, and J. Wu, "Detecting prefix hijackings in the internet with argus," in *IMC*. ACM, 2012, pp. 15–28.

[41] G. Siganos and M. Faloutsos, "Neighborhood Watch for Internet Routing: Can We Improve the Robustness of Internet Routing Today?" in *INFOCOM*. IEEE, 2007, pp. 1271–1279.

[42] O. Thonnard, "A multi-criteria clustering approach to support attack attribution in cyberspace." Ph.D. dissertation, École Doctorale d'Informatique, Télécommunications et Électronique de Paris, March 2010.

[43] O. Thonnard, L. Bilge, G. O'Gorman, S. Kiernan, and M. Lee, "Industrial espionage and targeted attacks: Understanding the characteristics of an escalating threat," in *RAID*. Springer, 2012, pp. 64–85.

[44] O. Thonnard and M. Dacier, "A strategic analysis of spam botnets operations," in *CEAS*. ACM, 2011, pp. 162–171.

[45] V. Torra, "The weighted OWA operator," *Int. Journal of Intelligent Systems*, vol. 12, no. 2, pp. 153–166, 1997.

[46] P.-A. Vervier, Q. Jacquemart, J. Schlamp, O. Thonnard, G. Carle, G. Urvoy-Keller, E. W. Biersack, and M. Dacier, "Malicious BGP Hijacks: Appearances Can Be Deceiving," in *International Conference on Communications (ICC) Communications and Information Systems Security (CISS) Symposium*. Sydney: IEEE, June 2014, pp. 884–889.

[47] P.-A. Vervier and O. Thonnard, "SpamTracer: How Stealthy Are Spammers?" in *5th International Traffic Monitoring and Analysis (TMA) Workshop (INFOCOM workshops)*. Turin: IEEE, April 2013, pp. 453–458.

[48] R. Yager, "On ordered weighted averaging aggregation operators in multicriteria decision-making," *IEEE Trans. Syst. Man Cybern.*, vol. 18, no. 1, pp. 183–190, 1988.

[49] Z. Zhang, Y. Zhang, Y. C. Hu, Z. M. Mao, and R. Bush, "iSPY: Detecting Ip Prefix Hijacking on My Own," in *SIGCOMM*. ACM, 2008, pp. 327–338.

[50] C. Zheng, L. Ji, D. Pei, J. Wang, and P. Francis, "A light-weight distributed scheme for detecting IP prefix hijacks in real-time," in *SIGCOMM*. ACM, 2007, pp. 277–288.

APPENDIX

MULTI-STAGE ANOMALY SCORING AND AGGREGATION

We use Multi-Criteria Decision Analysis (MCDA) to design a multi-stage decision-making process and identify the most interesting cases by ranking IP blocks according to their (anomalous) routing behavior. A typical MCDA problem consists to evaluate a set of alternatives w.r.t. different criteria using an *aggregation function* [17]. The outcome of this evaluation is a global score obtained with a well-defined aggregation model that incorporates a set of constraints reflecting the preferences and expectations of the decision-maker. An aggregation function is defined as a monotonically increasing function of n arguments ($n > 1$): $f_{aggr} : [0,1]^n \longrightarrow [0,1]$.

OWA extends averaging functions by combining two characteristics: (i) a weighting vector (like in a classical weighted mean), and (ii) *sorting* the inputs (usually in descending order), hence the name of *Ordered Weighted Averaging* [48]. OWA is defined as:

$$OWA_{\mathbf{w}}(\mathbf{x}) = \sum_{i=1}^{n} w_i x_{(i)} =< \mathbf{w}, \mathbf{x}_{\searrow} >$$

where $\mathbf{x}_{\searrow}$ is used to represent the vector $\mathbf{x}$ arranged in decreasing order: $x_{(1)} \geq x_{(2)} \geq \ldots \geq x_{(n)}$. This allows a decision-maker to design more complex decision modeling schemes, in which we can ensure that only a portion of criteria is satisfied without any preference on which exactly (e.g. "at least" k criteria satisfied out of n). OWA differs from a classical weighted means in that the weights are not associated with particular inputs, but rather with their *magnitude*, and it can thus emphasize the largest, smallest or mid-range values.

It might be useful also to take into account the *reliability* of each information source in the aggregation model, like in Weighted Mean (WM). Torra proposed thus a generalization of OWA, called *Weighted OWA* (WOWA) [45]. This aggregation function quantifies the *reliability* of the information sources with a vector $\mathbf{p}$ (as the weighted mean does), and at the same time, by weighting the values in relation to their relative *ordering* with a second vector $\mathbf{w}$ (as the OWA operator). *Weighted OWA* is defined by:

$$WOWA_{\mathbf{w},\mathbf{p}}(\mathbf{x}) = \sum_{i=1}^{n} u_i x_{(i)},$$

where $x_{(i)}$ is the i^{th} largest component of $\mathbf{x}$ and the weights u_i are defined as

$$u_i = G\left(\sum_{j \in H_i} p_j\right) - G\left(\sum_{j \in H_{i-1}} p_j\right)$$

where the set $H_i = \{j | x_j \geq x_i\}$ is the set of indices of the i largest elements of $\mathbf{x}$, and G is a monotone non-decreasing function that interpolates the points $(i/n, \sum_{j \leq i} w_j)$ together with the point $(0,0)$. Moreover, G is required to have the two following properties:

1. $G(i/n) = \sum_{j \leq i} w_j$, $i = 0, \ldots, n$;
2. G is linear if the points $(i/n, \sum_{j \leq i} w_j)$ lie on a straight line.

When the number of criteria to be evaluated is large, it is generally considered a best practise to organise them in subgroups, which are then evaluated hierarchically. Figure 10 illustrates the design of our multi-stage anomaly scoring and aggregation system, in which we organise the aggregation of anomalies in different subgroups based on their semantics. The advantage of a multi-stage aggregation model is that intermediate decision thresholds are not needed. Intermediate aggregate scores are propagated up to the highest level where they can contribute to the overall score.

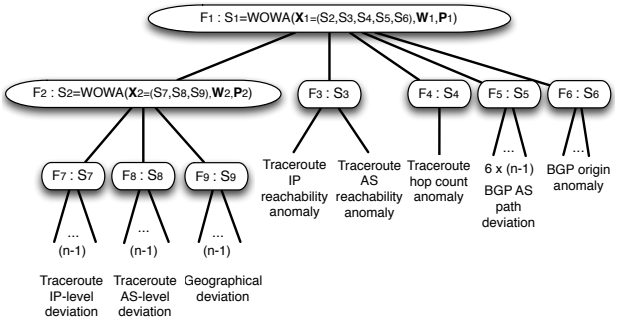

Fig. 10. Evaluation and ranking of candidate BGP hijacks: a multi-level aggregation model $\mathcal{M}$ for anomaly scores.

Given the definitions here above, we define an aggregation function $\mathcal{F}_a$, with as output the aggregated score S_a given by WOWA calculated for the anomaly a as:

$$\mathcal{F}_a : S_a = WOWA(\mathbf{x}_a, \mathbf{w}_a, \mathbf{p}_a)$$

where $\mathbf{x}_a$ is the vector of scores to aggregate and $\mathbf{w}_a$ and $\mathbf{p}_a$ the WOWA weighting vectors. As shown in Figure 10, we can then define our multi-stage anomaly scoring and aggregation model $\mathcal{M}$, with as output the final score S_1, for a given spam network, as the recursive function $\mathcal{F}_a$ where:

$$\begin{aligned}
&\mathcal{F}_1 : S_1 = WOWA(\mathbf{x}_1 = (S_2, S_3, S_4, S_5, S_6), \mathbf{w}_1, \mathbf{p}_1)\\
&\mathcal{F}_2 : S_2 = WOWA(\mathbf{x}_2 = (S_7, S_8, S_9), \mathbf{w}_2, \mathbf{p}_2)\\
&\quad \ldots\\
&\mathcal{F}_9 : S_2 = WOWA(\mathbf{x}_9 = (a_{geo_1}, \ldots, a_{geo_{n-1}}), \mathbf{w}_9, \mathbf{p}_9)
\end{aligned}$$

As an example, we define $\mathbf{w}_1 = (0.5, 0.3, 0.2, 0.0, 0.0)$ and $\mathbf{p}_1 = (0.2, 0.25, 0.15, 0.25, 0.15)$ to obtain the final score S_1 as outcome of the top-tier aggregation stage. Vector $\mathbf{w}_1$ translates here the intuition that a hijacked spam network does not always exhibit all anomalies (*e.g.,* a hijack does not necessarily involve a BGP origin anomaly) hence we require that "at least some" of the anomaly scores have a high score to contribute to a final aggregate score above a predefined decision threshold. The components of $\mathbf{p}_1$ translate the confidence we have in the different anomaly types to identify a suspicious routing change. The highest confidence score (0.25) is assigned to the traceroute reachability anomaly S_3 and BGP AS path deviation S_5, which by experience have proved being particularly reliable. On the other hand the traceroute hop count anomaly S_4 and BGP origin anomaly S_6 are assigned a lower confidence score (0.15) because we observed them only in a few rare hijack scenarios. Finally, the traceroute path deviation S_2 is given a medium confidence (0.2) as it can be affected by inaccuracies in traceroute measurements. The model parameter definition is done similarly at the other intermediary stages (for w_i, p_i where $i = 2, \ldots, 9$) so as to include expert knowledge and model the preferences of a network analyst.