Demystifying the IP Blackspace

Quentin Jacquemart[1] ✉, Pierre-Antoine Vervier[2],
Guillaume Urvoy-Keller[3], and Ernst Biersack[4]

[1] Eurecom, Sophia Antipolis
quentin.jacquemart@eurecom.fr
[2] Symantec Research Labs, Sophia Antipolis
Pierre-Antoine_Vervier@symantec.com
[3] Univ. Nice Sophia Antipolis, CNRS, I3S, UMR 7271, 06900 Sophia Antipolis
urvoy@unice.fr
[4] erbi@e-biersack.eu

Abstract. A small part of the IPv4 address space has still not been assigned for use to any organization. However, some of this IP space is announced through BGP, and is, therefore, globally reachable. These prefixes which are a subset of the *bogon* prefixes, constitute what we call the *blackspace*. It is generally admitted that the blackspace stands to be abused by anybody who wishes to carry out borderline and/or illegal activities without being traced.
The contribution of this paper is twofold. First, we propose a novel methodology to accurately identify the IP blackspace. Based on data collected over a period of seven months, we study the routing-level characteristics of these networks and identify some benign reasons why these networks are announced on the Internet. Second, we focus on the security threat associated with these networks by looking at their application-level footprint. We identify live IP addresses and leverage them to fingerprint services running in these networks. Using this data we uncover a large amount of spam and scam activities. Finally, we present a case study of confirmed fraudulent routing of IP blackspace.

1 Introduction

The global BGP (Boder Gateway Protocol) routing table now contains over 600k distinct IPv4 prefixes. A few of these prefixes should not be globally announced (such as the private IP space) and are collectively referred to as *bogon* prefixes. A subset of bogon prefixes, which we call the *blackspace*, is composed only of prefixes that have not been assigned for use to any organization.

These unallocated, yet globally announced and reachable blackspace prefixes traditionally hold a bad reputation. On top of uselessly cluttering up the global routing table, there have been reports of DDoS (Distributed Denial of Service) attacks originated from blackspace address blocks [19]. Spammers are also believed to abuse the blackspace in order to stealthily announce and abuse routes [8]. By extension, it is admitted that the blackspace stands to be abused by anybody who wishes to carry out borderline and/or illegal activities without being traced.

Because it is unallocated, hijacking a blackspace prefix is more likely to go unnoticed. Traditional hijacking detection tools, such as Argus [16], focus on “regular”

prefix hijackings, i.e. situations in which the hijacked prefix is announced by the attacker alongside the owner's legitimate announcement. In the case of blackspace prefixes, there is no rightful owner, and thus no legitimate announcement that can be used to find an anomaly. Consequently, hijacking blackspace prefixes is out of the detection scope of state-of-the-art monitoring tools. Hijacking a blackspace prefix is also different from hijacking a *dormant* prefix, as analyzed in [22]. Dormant prefixes have been handed out for active use to organizations, but are globally unannounced; whereas blackspace prefixes are unallocated, and *should not* be globally announced.

Therefore, it is recommended to filter out bogons (including the blackspace), so as to minimize the window of opportunity of potential abusers. Unfortunately, the blackspace constantly varies in size and shape, according to new prefix assignments and prefix returns that are carried out daily by different Internet actors. Filtering out bogons is therefore inconvenient and tricky. In order to automate the process as much as possible, Team Cymru provides multiple lists with different levels of granularity that can be included directly in a BGP router's configuration [18].

This paper focuses on the study of blackspace prefixes and aims to clarify what the blackspace contains. A partly similar study, which encompassed all bogon prefixes [8], was carried out over 10 years ago. The formal reporting of malicious events carried out from the blackspace, [19], is even older. Back then, the IPv4 landscape was much different from today's, and the results provided by these works are not applicable anymore in today's Internet.

We start by detailing the method that we use to isolate the blackspace prefixes from the BGP routing table. We then provide a thorough study of the blackspace networks on two different levels. First, we look at the information we extract from the BGP control plane and study the size of the blackspace. We then study the persistence and change in the blackspace through time. We characterize the origin ASes (Autonomous Systems) that actively announce blackspace by using semantic information (e.g. WHOIS records). Second, we look at the data plane and focus exclusively on the security threat associated with the blackspace prefixes. In order to do so, we actively seek live IP addresses and extract the domain name for these machines. We check the websites running in the blackspace, analyze their content, and check if their URLs are known to be malicious. We use an IP blacklist to locate hosts that are associated with adware, scam, phishing, and other malicious activities. Finally, we check for spamming activities and show how some spammers skillfully abuse the unallocated IP space in order to remain anonymous.

This paper is organized in the following way. Section 2 details the method and the datasets we use in order to locate the blackspace inside the BGP routing table. Section 3 details our analysis results: Section 3.1 studies the size and variation of the blackspace; Section 3.2 details the BGP topology characteristics of the blackspace prefixes; Section 3.3 details the active measurements we do on blackspace networks, as well as a detailed threat analysis. Section 4 discusses the shortcomings of our approach. Section 5 provides a summary of the existing work and this domain, and how our efforts differ, and improve the current state-of-the-art. Finally, Section 6 summarizes our findings and provides a few ways to improve our system.

2 Isolating the Blackspace

In this Section, we detail how we isolate the blackspace prefixes within the global BGP routing table by using a combination of distinct datasets that provide information about IP assignments. This step is necessary because there is no information on how the current bogon list [18] is populated. We show later, in Section 5, that our methodology for identifying the IP blackspace is more accurate and finer grained than previous efforts.

2.1 IP Space Assignation Hierarchy

To better understand our methodology, it is perhaps best to first briefly mention how the IP address space is divided into multiple blocks by distinct institutions before being assigned to end users, such as ISPs, corporations, or academic institutions. First, the IANA (Internet Assigned Numbers Authority) is in charge of distributing /8 prefixes to RIRs (Regional Internet Registries). There are five RIRs, each responsible for a different geographical area. In turn, RIRs allocate IP address space to LIRs (Local Internet Registries), such as ISPs, large corporations, academic institutions, etc. LIRs enforce their RIR's policies and distribute IP address blocks at the local level, i.e. to end users [1, 14].

2.2 Definitions

Bogon prefixes have traditionally been loosely defined as any IP prefix in the BGP routing table that should not be globally reachable. More precisely, following the definitions of [18], a prefix is a **bogon** if any of the three following conditions is true: i) it is a **martian** prefix, i.e. if it is a prefix that was reserved for special use by an RFC, such as the private IP address space; ii) the prefix belongs to a block that was not assigned to any RIR by the IANA; iii) the prefix belongs to a block that was not assigned by a RIR to a LIR, or to an end user.

We define the **blackspace** prefixes as the set of bogon prefixes that are not martians and that are announced in BGP. In other words, it is the set of BGP-announced prefixes that have not been assigned for use – either because it still belongs to the IANA pool, or because a RIR has not assigned it to an ISP or an end user. We explicitly remove martian prefixes because they are most likely the result of a local route leak caused by a misconfiguration [8]. Moreover, since these prefixes are internally routed in a lot of networks, we are unlikely to reach martian-originating networks from our own, rendering any standard network diagnostics utility such as `ping` or `traceroute` pointless.

2.3 Internet Routing Registries

The IRRs (Internet Routing Registries) are a set of distributed databases maintained by the five RIRs where network operators can provide information regarding their network on a voluntary basis. In particular, the `inetnum` objects contain information regarding IP address space assignment [2]. Consequently, the IRR databases sound like the ideal starting point to isolate the IP blackspace. We need to access the database of each RIR,

and extract the IP ranges mentioned in `inetnum` objects. We then have to check the prefixes announced in BGP against the ones we found in the IRRs, and keep those that do not match.

Unfortunately, things are not quite that simple. Like previously stated, providing information in the IRR databases is in no way mandatory, and even though it is considered as a good practice for LIRs to maintain their allocation information up to date, they are in no way required to do so. Additionally (and somehow consequently), the IRR databases are manually updated, and thus are plagued with typical human errors, such as typos. For example, some `inetnum` objects end their network on a .2**2**5 IP address, where the right value would be 2**5**5; some objects explicitly discard their net address, and/or their broadcast address, etc. Due to these reasons, we cannot expect to have an exact mapping between the BGP prefixes and the IRR prefixes. As a result, if we cannot match a BGP prefix to an IRR prefix, we take into consideration `inetnum` objects that are within the BGP prefix (i.e. `inetnum` objects that are more specific than the BGP prefix). If over 95% of the address space of the BGP prefix is covered by more specific IRR prefixes, we consider the BGP prefix has having been assigned, and that providing a matching IRR entry was overlooked. Our reasoning is that each customer of LIRs (which may be other ISPs) potentially wishes to update the IRR database, if only to update the management information of their network, such as technical and administrative contact details.

2.4 RIR Statistics Files

Every day, each RIR publishes a report – sometimes known as the delegation report – on the current status of the use they make of resources they have been allocated, including IP address space [3]. This report breaks down each RIR's IP address pool into four distinct states: ALLOCATED, ASSIGNED, AVAILABLE, and RESERVED. The first two states, ALLOCATED and ASSIGNED, are similar in the fact that they both have been marked as usable by someone by the RIR, i.e. these addresses can be announced. The difference is that ALLOCATED space ought to be used by LIRs for suballocation, whereas ASSIGNED space should not – i.e. it should be used directly by the LIR or end user. As the name suggests, the AVAILABLE state contains addresses that have not been ALLOCATED or ASSIGNED to any entity. Finally, the RESERVED state is somehow an intermediate between the other states: it has not been ALLOCATED (or ASSIGNED) to anybody, but is also not AVAILABLE for such purposes. For example, these addresses might be reserved for the growth of a LIR, returns that have not been cleared yet, or experimental space [3]. In this classification, the blackspace is shared between RESERVED and AVAILABLE states: in both cases there should not be any public BGP announcement for these addresses.

2.5 Blackspace Computation Process

Our BGP dataset is built on the data provided by the RIPE RIS collectors [15]. We daily fetch the routing table of each of the 13 active, geographically diverse routers, and create a list of all globally reachable routes. In the same time, we daily extract all `inetnum` objects from each IRR database, and we compare these two datasets as

described in Section 2.3. We then remove from the remaining BGP prefixes the parts for which there exists an IRR entry. For illustrative purposes, let's consider (a real-world case) where a /21 prefix is announced in BGP, and where only one of the /22 more specific prefixes has an `inetnum` entry. We remove the /22 that is in the IRR from the blackspace, leaving only the other /22 in it. At this point, there is a one-to-n relationship between the prefixes in the blackspace and the prefixes as announced in BGP: a single BGP-announced prefix can result in multiple entries in the blackspace once the registered parts have been removed.

We further filter the results by discarding prefixes that are marked as ASSIGNED or ALLOCATED by RIRs in their statistics files. Once more, there are cases in which the remaining prefixes are in multiple states wrt. the statistics files states, e.g. the IP space is ALLOCATED and RESERVED. In this situation, we only keep the part of address space that is either RESERVED or AVAILABLE.

It is noteworthy that, although using both the IRRs and the statistics files might appear redundant, there are documented inconsistencies between the two distinct datasets [10]. Because we aim at investigating the blackspace, it is essential to use these multiple sources in order to circumvent the limitations inherent to each dataset and to focus exclusively on real blackspace prefixes so as to avoid introducing bias in our results.

3 Blackspace Analysis

In this Section, we study the blackspace networks over a period of seven months, between September 2014 and March 2015. In Section 3.1 and Section 3.2, we consider the routing-level characteristics of the blackspace networks, and identify some patterns for legitimate blackspace announcements. Then, in Section 3.3, we seek to determine the security threat posed by the blackspace networks by looking at the application-level services running in these networks, and by checking whether they were involved in some malicious activities like spamming or scam website hosting. Finally we provide a case study of a confirmed case of cybercriminals who carried out nefarious activities such as spamming by abusing AVAILABLE IP space.

3.1 Prevalence and Persistence

In this Section, we focus on a few essential aspects of the blackspace by looking at the size, temporal characteristics, and variation of the blackspace. In order to observe those, we computed the blackspace once per day between September 1st, 2014 and March 31, 2015 with the method detailed in Section 2. We compute the blackspace once a day because the IRR databases we use and the RIR statistic files are updated with this same frequency.

During our observation, the number of globally distinct prefixes from our collector routers varied between 550k and 600k prefixes. These prefixes route around 180 equivalent /8 IP addresses, i.e. the equivalent of 180 class A networks, or 180×2^{24} IP addresses. The reason we focus on the number of IP addresses instead of the number of prefixes is that, because of the methodology explained in Section 2, the relationship

between a BGP prefix and a blackspace prefix is a one-to-many. By taking an aggregated BGP prefix and removing parts of it, we virtually inflate the number of prefixes in the blackspace, even though this larger number of prefixes actually represents a smaller IP space, rendering the prefix count meaningless. Figure 1 plots the daily number of IP addresses in the blackspace, as seen from a global BGP point of view. It shows that the blackspace size normally varies between between 10^{-2} and 10^{-1} eqv. /8. It also shows that this number is relatively stable, apart from two peaks in October 2014 and January 2015. We investigated the reasons behind these peaks and attributed them to the announcement of 192.0.0.0/2 between October 15, 2014 and October 20, 2014; and a series of smaller prefixes between January 24, 2015 and January 29, 2015. We classify these events as routing leaks because they meet the criterias behind BGP misconfigurations detailed in [11]: a relative short-duration, and low visibility. Only three collector routers received the a route for 192.0.0.0/2 in October, and only one received the multiple prefixes in January 2015. Moreover, in both cases, only a single Autonomous System path (AS path) was seen, and the origin AS was a private AS number. All in all, Figure 1 shows that the entirety of the blackspace could generally be contained in a single prefix, whose CIDR length would be between between a /10 and a /15.

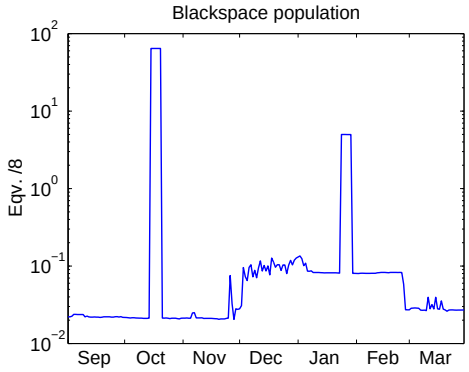

Fig. 1: Number of IP addresses in the blackspace, between September 1st, 2014 and March 31, 2015.

As mentioned in Section 2, a prefix in the blackspace has no `inetnum` entry in the IRR, and has not been allocated for use by a RIR. Figure 2 breaks down the statuses attributed to these IP addresses. Route leaks excluded, most of the blackspace is actually due to RESERVED resources, which are set aside by RIRs because they cannot be allocated right away.

Figure 3 plots the Cumulative Distribution Function (CDF) of the number of consecutive days a single prefix was included in the blackspace. The plain line plots this duration for all blackspace prefixes, including the many transient ones that were the results of the two route leaks already observed in Figure 1. The dashed line plots the same

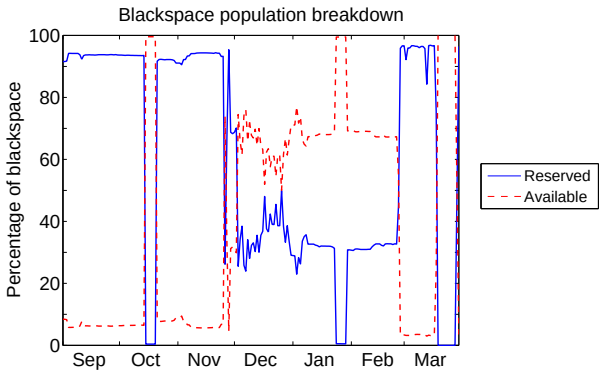

Fig. 2: Daily proportion of RESERVED and AVAILABLE address space in the blackspace, between September 1st, 2014 and March 31, 2015.

duration, but excludes the prefixes resulting from these leaks. The difference between these two curves implies that a lot of distinct prefixes were added to the blackspace due to the leak of routes. Indeed, the plain CDF shows that most blackspace prefixes are detected during four or five consecutive days, which is precisely the duration of the two leaks observed in Figure 1. On the other hand, the dotted CDF shows that 50% of blackspace prefixes that are not the result of these leaks are seen for at least 12 days, and that around 28% of them are seen during one day or less. In order to know how much the blackspace varies daily, Figure 4 plots the Jaccard index in-between two successive days. We compute the Jaccard index between days d and $d + 1$ as the ratio of the number of blackspace prefixes that are detected on both days, divided by the total number of distinct blackspace prefixes detected on day d and $d + 1$. A Jaccard index value of 1 indicates that the computed blackspaces for days d and $d + 1$ are identical. Conversely, a Jaccard index value of 0 indicates that the computed blackspaces for days d and $d + 1$ are 100% different. The closer to 1 the value is, the more similar the two blackspaces are. Once again, the variation is quite high when the route leaks start and finish, as shown by the full line; but there is not a lot of daily variation otherwise (as shown by the dashed curve).

The duration of a prefix in the blackspace (Figure 3) as well as the variation of the blackspace (Figure 4) imply that some prefixes leave the blackspace. This is possible if any of the three following conditions are met: i) the prefix is withdrawn from BGP; ii) an `inetnum` entry is added in the IRR; iii) the prefix is marked as ALLOCATED or ASSIGNED by a RIR. Figure 5 plots the distribution of each event for prefixes that exited the blackspace during our observation period. Again, the values are plotted for all entries, and also only for entries that were not the result of route leaks. In both situations, the most likely cause is that the prefix has been withdrawn. The second cause is the creation of an `inetnum` entry in an IRR database. If the IRR entry is more specific than the blackspace prefix, another (more) specific prefix will be included in the blackspace instead. Consequently, a bit less than 45% of prefixes leave the blackspace because the

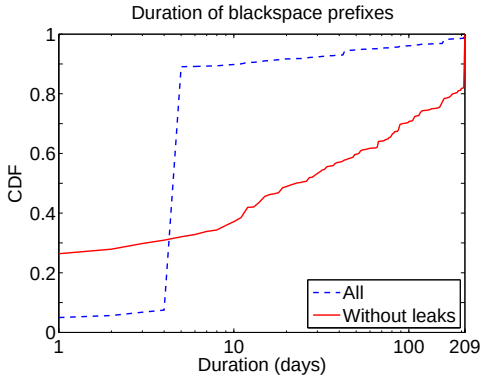

Fig. 3: Persistence of blackspace prefixes.

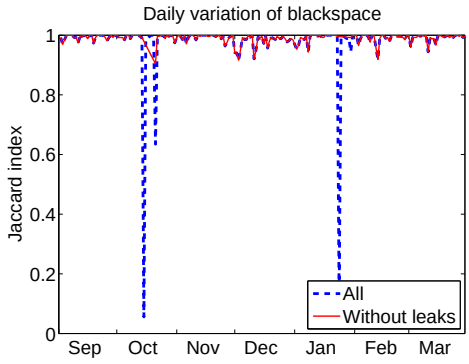

Fig. 4: Day-to-day variation of the blackspace prefixes.

BGP announcement was withdrawn. On the other hand, the other 55% become allocated (in one way or another) afterwards; which implies that half of the prefixes included in the blackspace are, potentially, used in good faith by the announcers. However, the other half, which globally amounts to a /11 network, does not end up as a registered network.

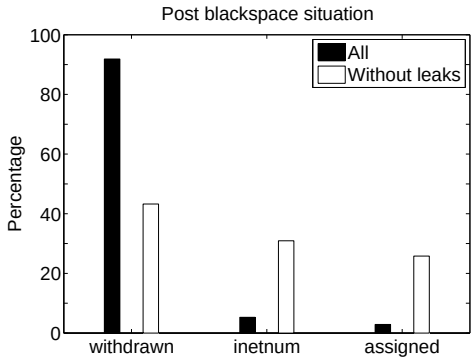

Fig. 5: Situation of the prefix after if left the blackspace.

3.2 BGP Characterization

In the previous Section, we saw that there are many blackspace prefixes, many of which are long-lasting. In this Section, we focus on the BGP characteristics of blackspace prefixes. We first focus on the origin AS of the blackspace prefixes to shed light on their uses. Where we cannot, we look at the temporal evolution of the blackspace prefix along with its origin AS in order to better understand the root cause.

AS numbers are assigned a status by RIRs, just like IP blocks (see Section 2): either ALLOCATED, ASSIGNED, AVAILABLE or RESERVED. Figure 6 plots the daily proportion of each AS status for ASes that originate a blackspace prefix. The plot has been further broken down by explicitly classifying the private AS numbers (between 64,512 and 65,535 [12]) separately from the RESERVED set. As can be seen by the black/squared line private ASNs are responsible for a large number of prefixes, but only during the two route leaks. In fact, all leaked prefixes are originated from a private ASN. ALLOCATED, ASSIGNED and RESERVED ASNs all roughly account for a third of blackspace prefixes, and AVAILABLE ASNs account for less than 10% of those. Just like with IP blocks, RESERVED and AVAILABLE ASNs are not ALLOCATED, and thus should not be in use. Yet, two thirds of the blackspace prefixes are originated by these ASes.

Figure 7 plots the percentage of blackspace prefixes for ASes that announce (at least) one blackspace prefix. The plot is further subdivided by AS status, but we excluded the private AS numbers, as they were the result of route leaks (see Figure 6). Here, both of the ALLOCATED and ASSIGNED statuses behave similarly, with more

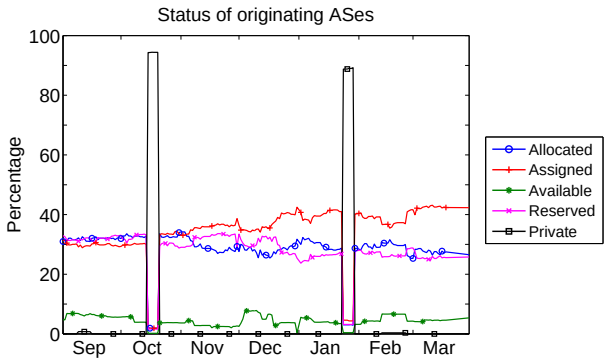

Fig. 6: Daily status of the ASNs originating a blackspace prefix.

than 90% of them announcing less than 1% of blackspace prefixes. Less than 10% of ALLOCATED (and around 20% of ASSIGNED) ASes originate more than a quarter of blackspace prefixes. On the other hand, close to 70% of RESERVED and AVAILABLE ASes *only* announce blackspace prefixes. To put this into perspective, the (global) average number of announced prefixes by ALLOCATED ASes is 229; by ASSIGNED ASes is 340; by RESERVED ASes is four, and by AVAILABLE ASes is two. In order to find out who operates these networks, we look at the names of the corporations behind these ASes (using [9]). We get 185 network names for ALLOCATED or ASSIGNED ASes that originate blackspace prefixes, for which we located the corporation website using mostly popular web search engines. We were able to resolve 178 names to mostly telephone or cable companies and ISPs (of all sizes and shape: tier-1 to tier-3, from dial-up to business-grade fiber providers, all around the world), hosting and cloud providers, data centers, IT service companies, and world-wide tech companies. Other companies operated as advertising, airlines, bank and insurances, constructions, courier and parcel delivery services, e-commerce, Internet exchange points, law firms, medical companies, military contractors, and online news. We could not resolve seven names. One was established as a company, but the website did not work, one used a name too generic to be found, and for three we could not locate any further information. The two remaining ASes appear to have been registered by individuals in Eastern Europe who also own other ASNs which are known to send spam – but do not originate blackspace prefixes at the same time.

Because the RESERVED and AVAILABLE ASes are not registered, we were not able to find registration information for them. Instead, we looked at the BGP topology of these prefixes, and investigated on the evolution of the blackspace prefix through time. For 33% of the cases where a blackspace prefix is originated from a RESERVED AS, the origin AS remains RESERVED throughout the whole observation period. The prefixes were marked as RESERVED. These networks are usually single-homed and peer either directly with a tier-1 provider, or with a tier-3. The other 66% prefixes show a state transition from, or to, RESERVED. In all the cases we observed, this was due to a network

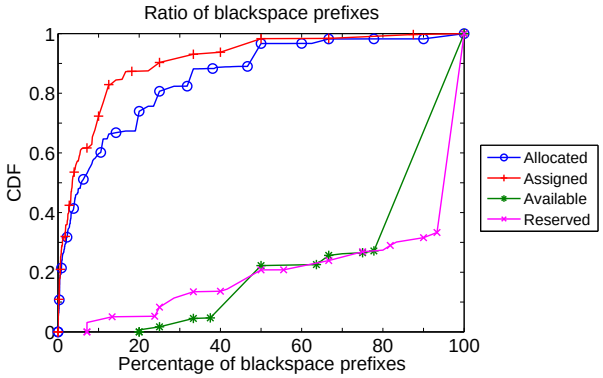

Fig. 7: Percentage of blackspace prefixes originated by ASes according to that AS's status.

owner either bringing up a new network, or decommissioning an old one. For example, half a dozen blackspace prefixes were originated from a RESERVED AS for six months through a tier-1 AS. On one day, the AS status changed to ASSIGNED and the name matched a well-known airline. The next day, the prefixes were all given inetnum entries in the IRR. Our interpretation is that the prefixes and ASN were RESERVED for the growth of said airline, and that they started using these resources before the paperwork had been fully processed. In another case, the prefixes and ASN were ALLOCATED, but one day turned to RESERVED. By looking up the company's name, we were able to find a letter from ICANN, informing the company that they had breached their registrar accreditation agreement by failing to meet technical requirements, and also by failing to pay the accreditation fees. The day following the date of the letter, all of that company's resource where changed to RESERVED. In some cases, there are transitions from ALLOCATED, to RESERVED, and then back to ALLOCATED. In this situation, we believe the situation was similar to the one of the last example, except that they corrected their behaviour to meet the requirements during the grace period. In the case of AVAILABLE ASes, there were only a handful of situations in which the AS (and the announced blackspace prefix) ended up as ALLOCATED or ASSIGNED. In these situations, it was the result of a new network being connected to the global Internet.

In conclusion, by looking at the routing-level characteristics, we were able to identify a set of blackspace prefixes that appear to be benign. Some prefixes appear to be in the blackspace because they have just been allocated, or because they are being phased out. Moreover, some blackspace networks are originated by tier-1 ISPs. Consequently, these networks are unlikely to be maliciously announced. All other networks need to be further analyzed in order to assess their threat level. To carry out this analysis, the next Section will be focusing on uncovering the application-level services running in the blackspace and seeking for hosts associated with malicious network activities.

3.3 Data plane and Application-Level Analysis

A. Introduction. In the previous Sections, we have explored the routing-level characteristics of blackspace networks. We have identified a small number of network practices leading to benign blackspace announcements. In order to be able to assess the security risk that is posed by the remaining set of blackspace prefixes, we need to know more about their network activities, e.g. which application-level services are running and whether they are known to be the source of some malicious network traffic. For this, we first need to find out live IP addresses and domain names, and we will then look at the services that these machines are running and check them against logs of malicious network activities. Table 1 summarizes our findings.

Domain-based reputation (Section 3.3.B)	Total	Domain names	556
		Hostnames	1,428
	Malicious	Domain names	35
		Hostnames	222
		IP addresses	142
		IP prefixes	81
	Benign	Domain names	5
IP-based reputation (Section 3.3.C)	Malicious	IP addresses	46
		IP prefixes	28
Spam (Section 3.3.D)	Malicious	IP addresses	206,404
		IP prefixes	58
SPAMTRACER [22] (Section 3.3.E)	Malicious	IP prefixes	82

Table 1: Breakdown of application-level activities in the blackspace.

In order to discover live IP addresses, we lightly probed each of the blackspace networks once per day in February and March 2015, except for ten days between Feb 16 and Feb 26 when our modem broke down. Using zmap [7], we sent a TCP SYN packet to each IP address included in a blackspace prefix on ports 21 (FTP), 22 (SSH), 25 (SMTP), 80 (HTTP), 137 (NetBios), 179 (BGP), and 443 (HTTPS). We run the scan from a machine located in AS3215 (Orange), and wait for SYN/ACK replies. Please note that the number of ports that we can scan is limited by the bandwidth we have been allowed to use for our experiments. The particular choice of the port number reflects what we believe to be the most popular services running on the Internet.

Figure 8 plots the number of SYN/ACK received per day and per port from the blackspace. There is quite a large number of web servers running in the blackspace. We customarily get replies from between 6k and 8k machines on port 80, and 2.5k machines on port 443. Next is port 22, with around 1k daily SYN/ACKs. There are around 100 FTP servers, and around 50 hits on port 179, suggesting suggesting that these IP addresses are border routers. Finally, we only get a handful of TCP replies on the NetBios port, and no reply at all on port 25.

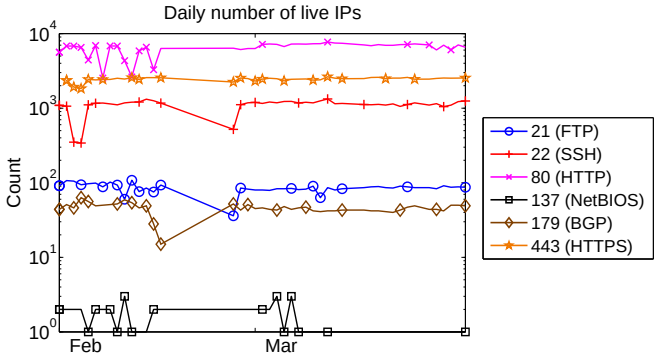

Fig. 8: Daily number of SYN/ACK packets received from the blackspace.

Figure 9 plots the variation of the live IP addresses in the blackspace, which indicates the persistence of these IP addresses. As we can see, the variation is quite high. These results need to be put into perspective of Figure 4 which showed that there was a very small variation in the blackspace networks. This suggests that the hosts inside blackspace networks are not static, but dynamically come and go. In other words, these networks appear to be actively configured, and not left in a 'legacy' state.

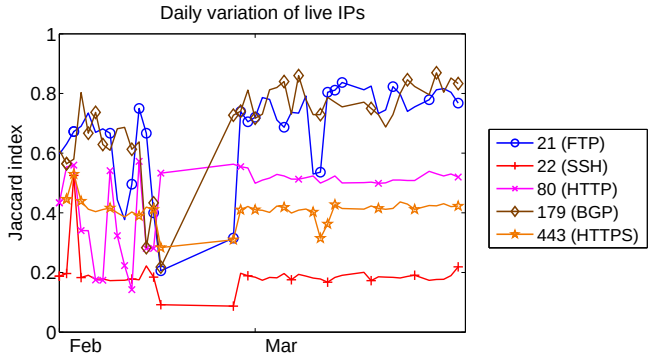

Fig. 9: Day-to-day variation of live IP addresses in the blackspace.

B. Websites, URLs, and Domain Names. In the previous Section, we located a set of highly volatile live IP addresses in the blackspace, and we saw that we found thousands of web servers daily. In this Section, we look at the contents of these websites and

their associated URLs and domain names which we match with a domain whitelist and blacklist. A simple way to know what's going on with these servers is to check the web page they serve. As a result, we supplement our scan with a simple HTTP client that just fetches the default page returned by the server, using the simple request `GET / HTTP/1.0`.

Using the returned HTTP headers, we find that over 90% of pages inside the blackspace are served by an Apache server; then come IIS, and Cisco IOS. Other pages are returned by nginx and lighthttpd, various application platforms, even including a print server. Because we get thousands of pages per day, we cannot manually go through all of them. In order to help our analysis, we used an unsupervised machine learning tool that clustered our pages based on the similarity of their raw content. We get between 60 and 80 clusters. The most important one contains over 4000 Apache error pages. This implies that, for the most part, the default page of web servers located in the blackspace is an Apache error page. Other clusters include default web pages of each HTTP daemon (e.g. your installation was successful). Websites hosted in the blackspace are usually in small clusters containing two or three IP addresses, which we checked in order to conclude that they represent wide variety of websites. A number of login pages are available, either to enter a configuration interface (e.g. a router/printer configuration login page), but also web applications such as Microsoft appliactions (e.g. Outlook Web Access), remote desktops (e.g. Citrix), content management systems, and other propietery corporate software. A large airline consistently served their default web page. We also found some SME businesses, such as technology firms (e.g. tier-3 ISPs and local shops), a second-hand car dealer, a law firm, and a private security company. Finally, a small number of clusters contained online forums. From the content of the topics available on the default page, their content varied from standard community interest (e.g. online gaming), to obvious copyright infriging file sharing boards. In some rare cases, the retrieved page contained a lot of obfuscated JavaScript code. We used wepawet [5, 25] to check it out, and it always remained benign.

We further extracted from a passive DNS database we maintain all fully qualified domain names (FQDNs) that resolved to an IP address within a blackspace IP prefix at the time the prefix was announced in BGP. We found a total 1,428 distinct FQDNs that accounted for 556 distinct domain names. We then checked these domain names against various blacklists including Spamhaus DBL [17] and VirusTotal [23][5] to search for scam, phishing or malware hosting activities associated with them. We also used the Alexa top 10,000 domain names as a means to determine whether some truly benign domains ended up being hosted on blackspace IP address space.

The correlation yielded 35 domains deemed malicious by the queried blacklists. These malicious domains were observed in no less than 222 different FQDNs, which appear to have resolved to 142 distinct IP addresses in 81 distinct blackspace IP address blocks. However, five domain names were also found in the Alexa top 10,000 ranked websites suggesting they were most probably benign. All of these were whitelisted, and belonged to well-known web applications, airlines, and technology companies. The remaining 516 domains could not be classified as either benign or malicious. From these observations we can see that while some blackspace announcements seem to be related

[5] VirusTotal includes more than 60 different website and domain scanning engines.

to legitimate activities, cybercriminals also appear to leverage such IP address space when performing nefarious activities.

C. Malicious IP Addresses. In order to locate host-level malicious activities inside blackspace prefixes, we were able to secure a list of malicious IP addresses from a IP-based reputation system that we maintain for operational purposes. These IP addresses were classified as either adware, phishing, scam, and other kinds of miscellaneous activity.

We looked for IP addresses that were included in blackspace prefixes exclusively on the days during which we detected the prefix in the blackspace. In other words, we explicitly discarded any matching IP address and its covering blackspace prefix where a match occurred outside of the blackspace period, even if there were matches during the blackspace period. The reasoning behind this (overly) strict matching is that we are looking for malicious activity that is the result of an individual abusing the blackspace in order to remain hidden. Thus, any matching malicious activity outside of the blackspace period could be argued to be the result of a previous owner of the prefix, and not from the blackspace itself. With these strict matches, we matched 46 malicious IP addresses in 28 distinct blackspace prefixes. Four of these IPs addresses were involved in scam activities, and the remaining 42 others in phishing activities.

We then looked into these eight BGP prefixes to see if we could obtain more information from the announcements. One of the BGP prefixes was RESERVED and originated by an AS that was marked as AVAILABLE, through what appears to be a tier-3 ISP in Thailand. Six of the other BGP prefixes were also all RESERVED, and originated by registered ASes. Two of these were country-wide ISPs, one was a television by satellite broadcaster, and one belonged to a hosting provider. A European prefix was being announced by the AS of a Japanese corporation, on which we were unable to find any information.

The remaining BGP prefix is 192.0.0.0/2, which we had previously classified as a route leak becaused it mached the descriptions in [11]. This prefix was announced between October 15, 2014 and October 20, 2014. This announcement resulted in an additional 2,970 prefixes in the blackspace (see Figure 1). Among these, 22 contain IP addresses marked as malicious, exactly during the announcement period. More precisely, a single /24, as well as a /19, both contain 11 individual malicious IP addresses, a /22 contains five, a /20 contains two. The remaining four IP addresses are spread across different blackspace prefixes. It is important to stress that the matches were done exclusively on the blackspace period. Actually, none of these prefixes were routable before or after this leak. The route also had a low visibility: it was only seen by 3 (out of 13) RIPE RIS collector routers; and there is only one single AS path leading to the origin. The origin AS was 65000, a private AS number (Figure 6), and the route was propagated through one cloud services and hosting provider, and then through a tier-3 ISP in the USA. Section 4 further discusses this peculiar situation.

D. Spam Campaigns. In an effort to further characterize the footprints of blackspace prefixes while they were announced and determine whether they pose a security threat to the Internet, we extracted spam source IP addresses in these prefixes that were

blacklisted in Spamhaus SBL and DROP (Don't Route Or Peer) [17], Uceprotect [21], PSBL [13] and WPBL [24]. Furthermore, we retained only those IP prefixes where spam activities were exclusively reported while the prefixes were announced as blackspace to ensure that the observed activities were not related to the previous or next status of the prefixes. We identified a total of 206,404 distinct spam sources in 58 IP prefixes. Figure 10 shows the BGP announcements and blacklisted spam sources related to a sample of 15 out of 58 blackspace prefixes while they were announced as blackspace.

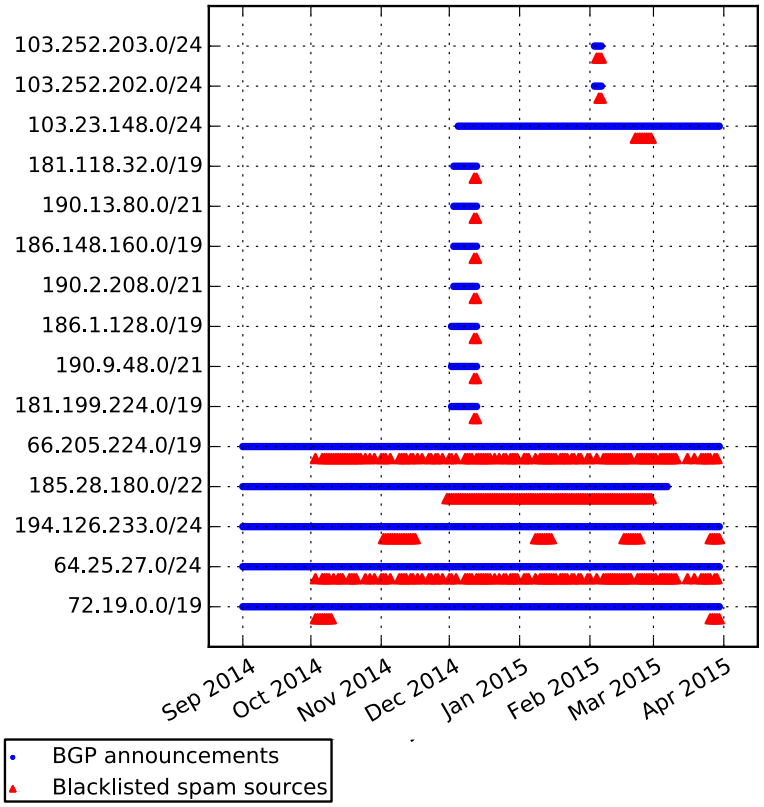

Fig. 10: BGP announcements and blacklisted spam sources related to IP prefixes while they were announced as blackspace. For the sake of conciseness, only 15 out of 58 prefixes that were blacklisted are depicted.

Finally, we correlated the list of blackspace IP prefixes with the output of SPAMTRACER [22], a system specifically designed to identify network IP address ranges that are hijacked by spammers to enable them to send spam while remaining hidden. Relying on a combination of BGP and traceroute data collected for networks seen originating spam and a set of specifically tailored heuristics, the system identifies those spam net-

works that exhibit a routing behavior likely indicating they were hijacked. We found that 82 IP prefixes were reported by SPAMTRACER as hijacked spam networks at the same time we identified them as being part of the blackspace.

E. Case Study. Starting from the 82 particularly suspicious blackspace prefixes we uncovered a very interesting phenomenon that we describe in-depth here below. Looking closely at how these 82 network prefixes were announced in BGP revealed that they were all advertised via one AS: AS59790 "H3S Helge Sczepanek trading as H3S medien services". Based on this intriguing observation, we decided to extract from all identified blackspace IP prefixes every of those that were advertised via AS59790. Surprisingly we discovered that no less than **476** IP prefixes in total (82 of them seen originating spam by SPAMTRACER) were advertised via AS59790 between October 17, 2014 and January 8, 2015 and that **all of them were part of the blackspace** at the time of the BGP announcements. Furthermore, all blackspace prefixes actually correspond to IP address space allocated by the IANA to AfriNIC (the African RIR) but not yet ALLOCATED or ASSIGNED by AfriNIC to any organization. Looking at the AS paths in the BGP announcements of the 476 networks

$$\{AS_{collector}, \ldots, AS174, \textbf{AS59790}\} \tag{1}$$

$$\{AS_{collector}, \ldots, AS174, \textbf{AS59790}, AS201509\} \tag{2}$$

reveals that AS59790 was always connected to a single upstream provider AS174 "Cogent Communications (US)", a cross-continent tier-1 ISP. From the AS paths we can also see that when AS59790 did not appear as the BGP origin AS (case 1) it was apparently used to provide transit to AS201509 (case 2). AS59790 "H3S Helge Sczepanek trading as H3S medien services (DE)" was ASSIGNED on September 30, 2014 and AS201509 "Sky Capital Investments Ltd. (DE)" was ASSIGNED on October 17, 2014, shortly before they started to be used to announce the blackspace prefixes. Both ASes were registered in the RIPE region to what appear to be organizations active in the finance industry in Germany. However, we were unable to find any information regarding these organizations through extensive web searches. The description of AS59790 and AS201509 in the IRR reveals that they are in fact under the control of the same person. We were unable to establish contact or to get any further information by contacting RIPE.

In summary,

- AS59790 and AS201509 were used to announce a total of 476 blackspace prefixes over a period of approximately three weeks;
- these ASes were never used to announced any non-blackspace prefix;
- some of the blackspace prefixes announced were used to send spam, according to [22].

The evidence presented here above suggests that these ASes were involved in malicious BGP announcements of IP blackspace. Moreover, a recent article from Dyn [6]

reported on similar evidence about AS59790 being involved in fraudulent routing announcements of unallocated African IP address space. This case study thus tends to confirm the assumption that blackspace IP prefixes are purposefully used to source different types of malicious network traffic, such as spam, likely in an effort to hinder traceability.

4 Discussion

In this Section, we address the shortcomings and weaknesses of our methodology.

The results presented in Section 3 offer a granularity of one day. This can be explained by the following reasons. First, the data sources that we use to compute the blackspace – i.e. the IRR databases and the RIR delegated files – are only updated once per day. Second, because we are actively probing the blackspace networks, we are effectively limited by the capacity of our Internet connection. In order to comfortably run this scan in its entirety (i.e. the equivalent /10 blackspace on 7 ports, with the additional web crawling), we need, on average, 17 hours. As a result, we cannot do more than a single scan per day. Third, and consequently, we use routing table dumps from RIPE RIS instead of BGP messages. Routing table dumps are generated every 8 hours and contain the entirety of the routes known by the router. The dumps of BGP messages are generated every 5 minutes and contain all the BGP messages exchanged between the collector routers and one of its peers. With those, we would obtain a much better granularity of data, maybe even include more prefixes in the blackspace. However, since we were mainly focusing on the accurate detection of blackspace prefixes, and on the discovery of the network footprints that they have, as well as the malicious activities they carry out, we think our results are still representative. Short-lived hijacks occurring in the blackspace would not enable an attacker to host a scam website, for example.

Our probing is done from a single machine located in AS3215 (Orange). While this gives us plenty of control over the environment in which our experiment is deployed, it comes at the price of a few drawbacks. First, we don't know anything regarding the BGP-view of the network we are connected in. In other words, we are using BGP data from RIPE RIS as the source of our control-plane data, and the Orange network in order to explore the connectivity. Even though Orange is a tier-1 network, we could not find any direct peering between 'our' AS and a RIPE collector. Actually, AS3215 is routed through AS5511 – better known as OpenTransit – which contains Orange's tier-1 infrastructure. This potentially leads to false negative in our measurements, especially in the case low-visibility prefixes, such as the route leak of 192.0.0.0/2 in which we detected malicious IP addresses (Section 3.3.C). Would probes sent from our vantage point have reached the originating network, or would they have been dropped because there would be no "route to host"? The optimal way to carry out these measurements is from a machine that runs BGP so as to assess the reachability of the destination.

At the beginning of Section 3, we saw two BGP events leading to a sudden and massive increase of the blackspace size. We classified these events as route leaks because they were only seen by a handful of RIPE collectors – three collectors for the leak in October; one collector for the one in January – and because there was only a single AS path between the collector(s) and the origin. However, because we also de-

tected malicious activities inside of them, the question of whether these events were deliberate attacks disguised as route leaks needs to be raised. Unfortunately, we cannot provide a definite answer. But a recent report underlined highly localised BGP hijacks, engineered to have a very low footprint, and to remain invisible from the point of views of route collectors [20].

5 Related Work

The oldest report of malicious activities carried out from the bogon address space dates back to 2001 with [19], where the author provided an analysis of the attacks carried out against an active web site. A large proportion of attacks originated from bogon addresses: 13% from within the bogons of classes A, B, and C; 53% from classes D (multicast) and E (future use). All in all, by properly filtering incoming traffic at a border router, 66% of attacks could easily be mitigated.

As a result, Team Cymru set up the bogon reference project [18], which precisely defines the different categories of bogon prefixes. We used this as the basis of our definitions in Section 2. Additionally, multiple lists of bogon prefixes are offered to network owners who wish to filter bogons out of their networks, which can be retrieved in many convenient ways and formats. These lists vary according to the desired level of precision. The bogon lists contain the prefixes still reserved in the IANA pool, as well as prefixes reserved by RFCs for specific use cases. The *full* bogon list supplements these prefixes with prefixes that have been allocated to RIRs by the IANA, but not by RIRs to ISPs or end users. These lists are dynamic, and network operators that use them should update their filters accordingly. Unfortunately, the methodology used to populate these lists is not disclosed. By comparing the full bogon list with our blackspace list, we were able to identify key differences. First, the full bogon list does not make use of the IRRs, as evidenced by many prefixes for which an `inetnum` object could be found. Second, the full bogon list appears to implement some heuristics based on the status of the prefixes. For example, we noticed that prefixes whose status transitioned from either ALLOCATED or ASSIGNED to RESERVED were not listed in the full bogon list. We also noticed that some prefixes that were RESERVED for a long time were not listed, although it might be that the transition happened before our data was gathered. We ignore the motivations behind these heuristics. However, the comparison of our blackspace list with the full bogon list on the same day shows that using the IRR databases in addition to the RIR delegation files improves the accuracy of the list.

In 2004, Feamster et al. [8] provided the first formal study of bogon prefixes by looking into the prevalence and persistence of bogon announcements, as well as the origin ASes leaking these prefixes. However, the authors did not explicitly focus on the blackspace, but rather on the equivalent of the (simple) bogon list. Consequently, 70% of the analyzed events actually involve the prefixes reserved for the private IP space. Only 40% of the events lasted longer than a day. In our analysis, this value is of 75% (Figure 3). The rest of the study cannot be directly mapped onto our results, even though the beginning of Section 3 provides results to similar questions. However, with the authors' methodology, there is a one-to-one mapping between the BGP routing table and the bogon analysis. With this, they can focus on the number of bogon prefixes

announced by an AS. In our case, we have a one-to-n relationship between the BGP prefix and the blackspace prefixes because we divide the BGP announcement in separate parts that may have been assigned independently. The authors also focus on the effect of bogon filtering and show that network operators who filter out bogon prefixes usually do not update their filters in timely fashion, resulting in reachability issues and potential denial of service. It is also worth noting that the bogon prefixes used for the study were composed of the 78 /8 prefixes that still belonged to the IANA pool back then (excluding class E). Today, the IANA pool only consists of one single /8 prefix, 0.0.0.0/8 (also excluding 240.0.0.0/4). As a result, the IP address space inside which our studies have been conducted is much different.

6 Conclusion

In this paper, we focused on the IP blackspace, which is composed of the set of prefixes that are globally announced through BGP but have not been assigned for use to any entity. We presented a thorough methodology to compute the blackspace by using a combination of data sources reflecting the current allocations of the IP space. We saw that the daily blackspace address space is equivalent to a /10 prefix, and that the prefixes that compose it change over time. We actively studied those networks from the BGP control plane point of view, and also from the data plane point of view. While we showed that some of the blackspace is composed of prefixes that are either being phased out of the Internet or being installed, a significant part of it does not result from normal network operations, such as assignments and decommissions. By cross-checking with various reliable security data sources, we were able to isolate malicious activities that only occurred during a period in which the monitored prefixes were inside the blackspace. Even by using our strict matching rules, and our limited, targeted view of these networks, the amount of malicious activities is significant. In particular, we showed through a validated case study that cybercriminals **do** abuse blackspace prefixes to carry out nefarious activities while also hindering their traceability.

Consequently, this paper confirms how important it is to precisely filter blackspace prefixes out of BGP. Because state-of-the-art hijacking detection tools (such as Argus [16]) do not focus on detecting this particular form of hijack, filtering out routes to the blackspace is the only active counter-measure that can be used today against blackspace hijacks. However, the shape of the blackspace is dynamic, and previous studies [8, 4] have illustrated that, when a bogon filter has been setup, it is obsolete because it is not updated, thereby affecting the connectivity towards networks that are being added to the Internet. Moreover, the current source of bogon filtering [18] does not take into account `inetnum` entries from IRR databases, thus including – and preventing access to – networks that have been assigned to a customer.

This paper also underlines the difficulty of using a ground truth in BGP. Even though the prefixes that we focused on all have in common the fact that they should not even be used on the public Internet, we were able to show cases where their use was the result of legitimate practices. As a result it is still quite difficult to automate the estimation of the danger resulting from a particular prefix in the blackspace.

We plan to improve on our system in the following ways. First, we plan to define a set of reliable heuristics that would discard benign blackspace announcements and only retain those that are potentially harmful, thus increasing the quality of filters installed on routers. Second, we would like to supplement our probing system with a traceroute infrastructure that would enable us to geographically locate the origin of these networks, and the diversity of their connectivity. This would enable us to see if there are specific parts of the networks that hijackers prefer to abuse. Third, we need to view the BGP control plane, as well as the data plane from the same vantage point in order to make sure we reach low visibility routes. For this, we need a set of geographically diversified machines that run BGP – each connected to a different set of peers – and from which we can run our measurement experiments. If this can be achieved, a bonus point would be to make the system run in real time, by detecting and probing networks as they come and go in the BGP routing table. Our results currently focus on the IPv4 address space, inside of which the unallocated space is getting smaller every day. It would be interesting to do the same measurements with IPv6, and see how the results compare. As a final remark, note that we are able to provide interested parties with more detailed results and to discuss future work that can be undertaken with this dataset and infrastructure.

References

1. APNIC: Understanding address management hierarchy. http://www.apnic.net/services/manage-resources/address-management-objectives/management-hierarchy
2. APNIC: Using Whois: Quick Beginners Guide. http://www.apnic.net/apnic-info/whois_search/using-whois/guide
3. ARIN: Extended Allocation and Assignment Report for RIRs. https://www.arin.net/knowledge/statistics/nro_extended_stats_format.pdf
4. Bush, R., Hiebert, J., Maennel, O., Roughan, M., Uhlig, S.: Testing the reachability of (new) address space. In: Proceedings of the 2007 SIGCOMM Workshop on Internet Network Management. pp. 236–241. INM '07 (2007)
5. Cova, M., Kruegel, C., Vigna, G.: Detection and Analysis of Drive-by-Download Attacks and Malicious JavaScript Code. In: Proceedings of the World Wide Web Conference (WWW) (2010)
6. Doug Madory: The Vast World of Fraudulent Routing. http://research.dyn.com/2015/01/vast-world-of-fraudulent-routing/ (retrieved on June 5, 2015) (January 2015)
7. Durumeric, Z., Wustrow, E., Halderman, J.A.: ZMap: Fast Internet-wide scanning and its security applications. In: Proceedings of the 22nd USENIX Security Symposium (Aug 2013)
8. Feamster, N., Jung, J., Balakrishnan, H.: An empirical study of "bogon" route advertisements. Computer Communication Review 35(1), 63–70 (2004)
9. Huston, G.: AS names. http://bgp.potaroo.net/cidr/autnums.html
10. Huston, G.: RIR Resource Allocation Data Inconsistencies. http://www.cidr-report.org/bogons/rir-data.html
11. Mahajan, R., Wetherall, D., Anderson, T.: Understanding BGP misconfiguration. SIGCOMM Comput. Commun. Rev. 32(4), 3–16 (Aug 2002)
12. Mitchell, J.: Autonomous System (AS) Reservation for Private Use. RFC 6996 (Jul 2013)
13. Passive Spam Block List: http://psbl.org/
14. RIPE NCC: FAQ: Becoming a member. https://www.ripe.net/lir-services/member-support/info/faqs/faq-joining
15. RIPE NCC: Routing Information Service. http://www.ripe.net/ris/

16. Shi, X., Xiang, Y., Wang, Z., Yin, X., Wu, J.: Detecting prefix hijackings in the internet with argus. In: Proceedings of the 12th ACM SIGCOMM Internet Measurement Conference, IMC 2012. pp. 15–28 (2012)
17. Spamhaus: http://www.spamhaus.org/
18. Team Cymru: The Bogon Reference. http://www.team-cymru.org/bogon-reference.html
19. Thomas, R.: 60 Days of Basic Naughtiness: Probes and Attacks Endured by an Active Web Site. http://www.team-cymru.org/documents/60Days.ppt (March 2001)
20. Toonk, A.: Recent BGP routing incidents - malicious or not. Presentation at NANOG 63 (Feb 2015)
21. Uceprotect: http://www.uceprotect.net/
22. Vervier, P.A., Thonnard, O., Dacier, M.: Mind your blocks: On the stealthiness of malicious BGP hijacks. In: NDSS 2015, Network and Distributed System Security Symposium (02 2015)
23. VirusTotal: https://www.virustotal.com/
24. Weighted Private Block List: http://www.wpbl.info/
25. Wepawet. http://wepawet.cs.ucsb.edu