The parallel lives of Autonomous Systems: ASN Allocations vs. BGP

Eugenio Nerio Nemmi
nemmi@di.uniroma1.it
Sapienza University of Rome

Francesco Sassi
sassi@di.uniroma1.it
Sapienza University of Rome

Massimo La Morgia
lamorgia@di.uniroma1.it
Sapienza University of Rome

Cecilia Testart
ctestart@csail.mit.edu
MIT

Alessandro Mei
mei@di.uniroma1.it
Sapienza University of Rome

Alberto Dainotti
alberto@caida.org
CAIDA, UC San Diego
Georgia Institute of Technology

ABSTRACT

Autonomous Systems (ASes) exist in two dimensions on the Internet: the administrative and the operational one. Regional Internet Registries (RIRs) rule the former, while BGP the latter. In this work, we reconstruct the lives of the ASes on both dimensions, performing a joint analysis that covers 17 years of data. For the administrative dimension, we leverage delegation files published by RIRs to report the daily status of Internet resources they allocate. For the operational dimension, we characterize the temporal activity of ASNs in the Internet control plane using BGP data collected by the RouteViews and RIPE RIS projects. We present a methodology to extract insights about AS life cycles, including dealing with pitfalls affecting authoritative public datasets. We then perform a joint analysis to establish the relationship (or lack of) between these two dimensions for all allocated ASNs and all ASNs visible in BGP. We characterize the usual behaviors, specific differences between RIRs and historical resources, as well as measure the discrepancies between the two "parallel" lives. We find discrepancies and misalignment that reveal useful insights, and we highlight through examples the potential of this new lens to help pinpoint malicious BGP activity and various types of misconfigurations. This study illuminates a largely unexplored aspect of the Internet global routing system and provides methods and data to support broader studies that relate to security, policy, and network management.

CCS CONCEPTS

• **Networks** → **Network measurement**.

ACM Reference Format:
Eugenio Nerio Nemmi, Francesco Sassi, Massimo La Morgia, Cecilia Testart, Alessandro Mei, and Alberto Dainotti. 2021. The parallel lives of Autonomous Systems: ASN Allocations vs. BGP. In *ACM Internet Measurement Conference (IMC '21), November 2–4, 2021, Virtual Event, USA.* ACM, New York, NY, USA, 19 pages. https://doi.org/10.1145/3487552.3487838

IMC '21, November 2-4, 2021, Virtual Event

ACM ISBN 978-1-4503-9129-0/21/11…$15.00
https://doi.org/10.1145/3487552.3487838

1 INTRODUCTION

The Internet is a network of independent networks called Autonomous Systems (ASes) that use the Border Gateway Protocol (BGP) [60] to exchange reachability information and effectively interconnect. The number of ASes operating on the Internet has been steadily increasing since its inception, with currently some 70 thousand ASes exchanging routing information in BGP. Autonomous systems are uniquely identified in BGP by their AS number (ASN), which is delegated to ASes by Regional Internet Registries (RIRs).

The link between a given network and the ASN it uses on BGP is key to the proper functioning of the routing infrastructure. However, other than common practices [35, 36, 50] and anecdotal evidence of abuses [33, 34, 73], little is known about the actual relation between the administrative delegation of an AS number and its related announcements in BGP. In this paper, we develop and apply an analysis methodology to investigate this relation in terms of the actual behaviors observed *in the wild* and extract novel insights.

We perform the first joint longitudinal analysis of ASN delegation records and ASNs' BGP activity. To this end, we restore and build datasets—over a 17-years time frame—that we use as a dual-lens to examine the life cycle of ASNs. We show that this combined perspective can reveal insight into various operational phenomena impacting the security and stability of inter-domain routing—including malicious behavior, misconfiguration, administrative delays, and failed deployments—and potentially inform discussion on best practices and policy.

Our key contributions are:

- We propose a method enabling a novel bi-dimensional lens to look at BGP activity across time, which puts into focus important behaviors by RIRs, operators, and malicious actors.
- We carry out a meticulous restoration of 17 years of delegation files from all five RIRs, learning about errors and inconsistencies present in this precious public source of data. We make available the restored data (on top of which we build our datasets).
- We perform a longitudinal analysis comparing per-RIR behavior and highlighting historical and present trends related to infrastructural growth and (re-)allocation policies.
- Through a taxonomization based on our joint (admin-operational) perspective, we perform an in-depth analysis of the life of Autonomous System numbers. Our analysis reveals a long list of patterns and behaviors that improve our understanding of current practices and anomalies and can inform the discussion around

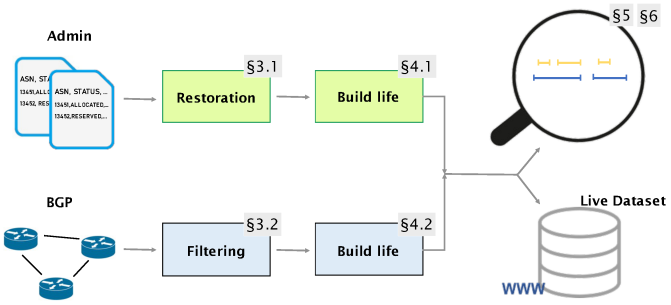

Figure 1: Representation of the pipeline of our work and the workflow of the paper.

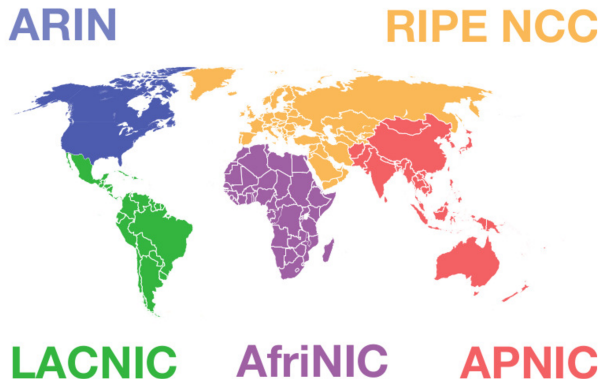

Figure 2: The five Regional Internet Registries (RIRs) manage Internet number resources in their respective regions of competence.

policy and best practices. Although in this work we do not develop a specific detection methodology, our results highlight the potential and practical relevance of ASN delegation data for identifying misconfigurations and malicious behavior.

- We publish our code and datasets for other works to leverage data on the administrative and operational lifetimes of ASNs in the Internet.[1] We will continue updating and publishing our datasets in order to facilitate near-realtime analysis and insight.

Roadmap. The diagram in Figure 1 illustrates the pipeline of this work. After providing background on AS number assignments in §2, in §3 we describe the ASN delegation and BGP datasets we use and our data sanitization and cleaning methods. In particular, we undertake a careful—and to the best of our knowledge, unprecedented—effort to verify and improve the consistency of the data provided in RIR delegation records in order to support our longitudinal analysis. In §4 we describe the methodology we use to build administrative and operational lifetimes out of these data. In §5 we present a first analysis of what we can learn by jointly looking at the administrative and operational dimensions at a broad (RIR-wide) scale. In §6 we delve into an in-depth joint analysis of the parallel lives of ASNs, highlighting insights about usual ASN behaviors, operational practices, inconsistencies, malicious activities, and misconfigurations.

2 AUTONOMOUS SYSTEMS AND THE INTERNET

From the moment the Internet became large enough to have "separate domains" in the early '80s, Autonomous Systems (ASes) needed to be identified by a specific number in routing protocols [47, 66]. Even though there is no verification step included in these routing protocols, the management of allocations of AS numbers and other Internet resources is required for the operation of the Internet [13]. From the first Autonomous System Number (ASN) delegation in 1983 [14] until now, the management and delegation of ASN has undergone substantial changes.

The early years. In the '80s, Jon Postel and the Internet Registry function of the Internet Assigned Numbers Authority (IANA) kept track of the assignations of AS numbers in RFCs [42, 58]. By 1990, 612 AS numbers had already been delegated. In the early '90s, following a recommendation by the IETF, the first Regional Internet Registries (RIRs) were created to manage Internet number resources—including AS number delegations—at a regional level, and leaving the IANA as the ultimate central authority, delegating large blocks of resources to the RIRs as needed [41]. Only in the early 2000s, RIPE NCC, ARIN, APNIC, and LACNIC, the registries for Europe, North America, Asia-Pacific, and Latin America and the Caribbean regions respectively, did start periodically publishing and archiving files with the status of Internet resource delegations. AfriNIC, the RIR for Africa, followed shortly after. Figure 2 depicts the different geographic areas covered by each RIR.

The initial daily tracking. While originally each RIR had its own format for keeping track of Internet number resource allocations in files—providing different information and published with different frequency—in 2004, the RIRs [2] unified the format and content [24] of the daily *"delegations files"*. Table 1 lists the dates of the first delegation file for each RIR. These files include information about AS numbers delegated, the registry that made the delegation, the country code of the organization to which the resource was allocated, and the date of the allocation.

The current delegation tracking. Between 2008 and 2010, the RIRs started using a new, *"extended"*, Internet resources delegation file format [25] initially developed by APNIC. This new format lists all the resources that are in the pool of each registry, including *(i)* the *available* resources that each RIR has—*i.e.,* resources that have been delegated by the IANA to each RIR to then allocate to organizations in its region—and *(ii) reserved* resources, which are resources in-between states: before either being delegated or returning to the pool of available resources. In addition, the extended format includes an opaque identification value in each line, the `Opaque_id`, to identify an organization within a file, so that resources allocated to the same organization all share the same Opaque_id. This new format provides a comprehensive picture of all the resources each RIR is responsible for and their respective status. There should be no overlap in resources between delegation files from different registries. All the registries but ARIN produce both the standard and the extended delegation files.

[1] Datasets and code available at https://github.com/SystemsLab-Sapienza/ParallelLives.

[2] At that time the RIRs were APNIC, ARIN, LACNIC, and RIPE NCC. AfriNIC was recognized as an RIR only in April 2005 [3].

The administrative life of an AS. The administrative life of an AS starts when a registry allocates a specific AS number to the given organization, removing that number resource from the available pool. The ASN will appear in the (extended) delegation file as *allocated*, with a corresponding registration date. The end of the administrative life happens when an ASN is either returned by the holder organization or reclaimed by the respective RIR, in accordance with RIR internal resources allocation policies. The ASN is then quarantined for some time in reserved status before going back to the available pool and being allocated again.

RIR-specific ASN allocation policies and reporting practices. RIRs have different approaches to handle ASN allocations, the eligibility criteria, the recovery of unused resources, the reuse of resources, and special cases (*e.g.,* ASNs reclaimed for a short time or ASN transfers), which impact ASNs' administrative lives. Appendix B describes in more detail the policies and how they have changed over time. For instance, since 2010, ARIN has been requesting number resources back from organizations that are out of compliance (*e.g.,* did not pay the annual fee), whereas other RIRs only actively reclaim unused resources or just reuse the ones given back to them or when the organization holding an ASN ceases to exist [11]. In addition, tracking in delegation files varies between RIRs for certain cases. For example, if an ASN held by a company is switched from *allocated* to *reserved*, and then it is allocated again to the same company, all RIRs except AfriNIC keep the registration date from the first allocation. Moreover, RIPE NCC and APNIC, do not modify the registration date of an ASN when it is transferred internally (inside the registry). Finally, APNIC allocates ASNs also to NIRs (National Internet Registries), thus introducing more uncertainty to when the NIRs allocate these resources to the end-users.

3 DATA COLLECTION & PREPARATION

This section describes our process to collect, restore, and sanitize the delegated files and BGP data we use in this study.

3.1 Restoring 17 years of ASN delegations

We collect all (regular/extended) delegation files from the RIRs' FTP sites [2, 5, 7, 45, 63], from the first file available (see Table 1 for details), until Mar 1, 2021; the RIRs FTP sites are publicly accessible. Across all RIRs, in less than 1% of the days in our observation time frame it happens that a (regular/extended) delegation file is missing from the site or the available file is corrupted. The longest count of consecutive days missing delegation files is 7 (RIPE). When both regular and extended delegation files are available[3] for the same day, we consider the information from the extended delegation file. The last column of Table 1 lists the total number of files collected per RIR, spanning a period of more than 17 years.

To be able to study the administrative lifetime of ASes through the lens of delegation files, we try to restore missing or potentially corrupted information. We make the restored data publicly available.[4] Our restoration process consists of the following steps.
***(i)* Filling the gap of missing files:** If an AS appears in both the day before and the day after an empty or missing file (157 occurrences), we assume that the AS is also allocated in the missing day. Otherwise, we use as reference for its starting (ending) date, the first (last) day it shows in the delegated files.
***(ii)* Filling missing records:** When comparing consecutive files, we find instances of large ASN count drops, although normally, the count monotonically increases. After careful investigation of large decrements, we find that in most cases when a group of ASes (from few hundreds to few thousands) disappears for one or a few days from the extended delegation file(s), we can recover information by leveraging the data still present in the corresponding regular delegation file(s).
***(iii)* Same day file update:** When comparing extended and regular delegation files from the same day, we find differences in 1.8% of the days—this happens for all RIRs except AfriNIC. We use the newest of the delegation files (based on the start and end times in the headers) to interpret the status of the ASNs accordingly. However, when an ASN disappears from the newest files for a few days but is *always* in the (corresponding) older files, we consider the ASN information in the old ones.
***(iv)* Cleaning invalid duplicate records:** In the AfriNIC files, we find duplicate records with inconsistent information (*e.g.,* allocated and reserved) persisting over periods of up to 6 months, with 16 ASNs affected in total. By manually looking at the history of each ASN, and sometimes their BGP behavior, we gather strong evidence disambiguating the inconsistent information.
***(v)* Restoring registration dates:** Some ASN delegation records show inconsistent registration dates, such as a registration date that is in the future with respect to the file date, that travels back in time across files, or that is filled with a placeholder value. We examine carefully each phenomenon and recover the registration date with the earliest date found in files when possible. For example, we find a few records in AfriNIC files for which the registration date is in the future when compared with the file date. As the difference is of a few days only, we use the date the ASNs first appeared in the delegation files (*i.e.,* the file date) as registration date.

We also find ASN delegations for which the registration date travels back in time (when only forward changes are expected, *i.e.,* new allocation). This type of phenomenon affects only few records in all RIRs except RIPE NCC, where more than 800 go backward in time to what we find is a "placeholder" registration date (1993-09-01). Most of these ASes are old ASes delegated in the '90s before the creation of most RIRs. Upon further inspection of these ASNs and contacting the respective RIRs, we trace back and confirm that these ASN allocations are all related to the ERX project: "early registration" ASN transfers from ARIN to the other

Table 1: Overview of the delegation files we collected from their inception until March 1, 2021 (between 16 and 17 years of data per RIR).

RIR	First regular	First extended	Number of files
AfriNIC	2005-02-18	2012-10-02	5,791
APNIC	2003-10-09	2008-02-14	6,345
ARIN	2003-11-20	2013-03-05	6,303
LACNIC	2004-01-01	2012-06-28	6,257
RIPE NCC	2003-11-26	2010-04-22	6,249

[3] Only ARIN completely stopped publishing the delegated files after Aug. 12, 2013; all the other registries decided to keep publishing both file types.
[4] https://github.com/SystemsLab-Sapienza/ParallelLives

RIRs [6, 8, 64, 65]. ARIN was formed in December 1997, and it inherited the database of existing address-block and ASN resources from InterNIC. In 2002, the RIRs agreed to have ARIN transfer the management of these resources to the respective RIRs accordingly to the region in which the holder of the resource resided. As a result, 5,026 ASNs were moved to APNIC, LACNIC, and RIPE NCC. We recover and restore the original registration dates leveraging delegation information published by ARIN before the delegation files era [8]. In a second phase of the ERX project, in 2005, once AfriNIC was created, it received 204 ASes in total from ARIN and RIPE NCC. However, in this case, the transfer did not alter the original registration dates.

***(vi)* Cleaning inter-RIR inconsistencies:** We find some 450 ASNs that—at different points in time—are simultaneously being allocated or reserved in multiple RIRs. We identify various overlaps, some affecting many ASes at once and lasting more than 250 days. After careful investigation, we find that the two main reasons for the multiple allocation of the same resources among RIRs are: *(i)* (regular or ERX) transfers where the "origin" RIR temporarily maintains stale data for ASNs that fails to remove from its delegation files and *(ii)* mistaken (apparent) allocations, some by RIRs who have not been assigned those ASN blocks from IANA. In all these cases, we are able to identify the cause and remove the evidently erroneous records from our data.

3.2 17 years of BGP data

To find operational ASN activity, we process historical BGP data from all available RIPE RIS [54] and RouteViews [68] collectors, using CAIDA BGPStream's Python library [57], starting on October 9, 2003 and ending on March 1, 2021. To track ASNs that appear in BGP paths, for each day, we process one full RIB dump per collector and *all* update dumps available.

Sanitizing BGP data: We sanitize the data discarding all paths to prefixes either longer than /24 or shorter than /8 for IPv4 and longer than /64 or shorter than /8 for IPv6, since they should not be globally propagated (except for specific cases such as *e.g.,* DDoS protection with BGP blackholing [20]). We also discard paths with loops since they are often related to misconfigurations [38]. A challenge when looking for all ASNs active in BGP is to distinguish low visibility ASNs from ASNs appearing because of errors in the BGP announcements a peer might share with a collector. In our long observation period the probability to incur into spurious data from 1 collector's peer is high. For this reason, we only consider an ASN to be active in BGP in a given day if in that day its visibility is strictly more than 1 peer, *i.e.,* two or more distinct ASes that peer with the collector infrastructure share BGP announcements with that ASN in the path that day.

In total we process **more than 930 billions RIB dump records and 2.3 trillion updates** over 17 years of data. We find a total of 96,391 unique ASNs being routed in BGP in the 17 years of our dataset, from 16,234 on October 9, 2003 to 73,143 on March 1, 2021.

4 BUILDING LENSES FOR ASN LIFETIMES

This section describes our methodology to build ASN lifetimes in terms of administrative allocations (§4.1) and BGP operations (§4.2). We show a snippet of the datasets in Listing 1. We make the datasets resulting from this process also publicly available, together with the code to generate them.[5]

4.1 Inferring ASN allocation lifetimes

Our method to infer *administrative ASN lifetimes* is based on two key fields in the delegations files—the allocation status and the registration date—in addition to the policies and practices followed by RIRs. As a general rule, we consider as the start of a new lifetime of an ASN the date of when it first appears or reappears (after deallocation) in the delegated files—or, in the case of extended delegated files, when it is labeled as *allocated*.[6] However, APNIC can allocate AS numbers in blocks to each of its National Internet Registries (NIR), which in turn allocate these resources to end-users. This characteristic introduces more uncertainty over the start of the actual administrative life.

We consider the end of a lifetime when it either becomes *available*, *reserved* or it disappears from delegation files. Specifically, we apply the following rules, which take into account different policies adopted by RIRs, either as documented or based on what we have learned in private conversations:

- ASN appearing allocated after being in *reserved* status or disappeared from the file.
  - An ASN is moved to the *reserved* status (extended delegated files) either if there are administrative issues with the organization that is holding the ASN or for quarantine, before the ASN becomes ready to be reallocated. We use as discriminating factor the registration date: if the ASN returns in the delegated files with the same registration date, it means it was not returned to the free pool, so we can assume it was returned to the previous owner and we merge the two allocation spans in one. Otherwise we infer it was reallocated to someone else.
  - Similarly, in the case in which the ASN disappears from the delegated files (when only regular delegated files are present), we consider the registration date the discriminant between reallocation (new date) and same owner/life (same date).
  - AfriNIC exception: for AfriNIC, if an ASN has been reserved for any period of time and becomes allocated without first being available, it means they re-allocated the resource to the previous owner even if it gets a new registration date. In this case, we merge the two allocation spans.
- Allocated ASN suddenly changing registration date: An ASN cannot be reallocated before being in quarantine. Thus changes in registration dates without ASNs being deallocated, are explainable by administrative corrections to the same current allocation.
- Inter-RIR transfers (342 in total): if an ASN is transferred across two RIRs, we consider the ASN allocation only one lifetime *iff* there are no gaps between the allocation in each RIR.

By applying these criteria, we identify 126,953 lifetimes, for a total of 106,873 ASNs, that have existed throughout our 17-year time frame of analysis.

[5] https://github.com/SystemsLab-Sapienza/ParallelLives

[6] Typically, this date is close to the registration date. Between 90.1% (AfriNIC) and 99.35% (ARIN) of the cases, the ASN appears in the delegation files the same day or the day after its registration.

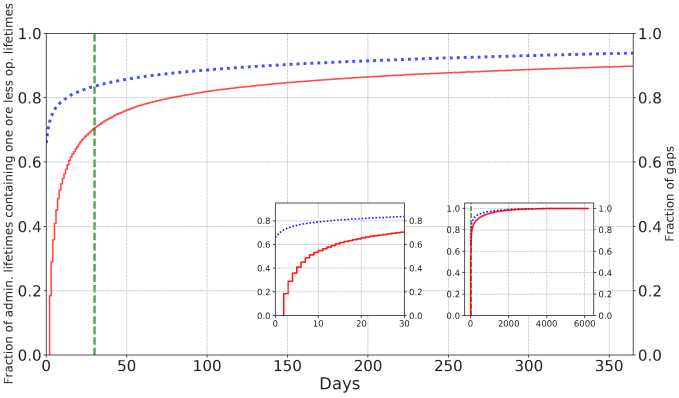

Figure 3: Sensitivity to different BGP activity timeout values: Distribution of per-ASN BGP activity gaps (red line) and fraction of administrative lives that contain one or no operational life (blue dotted line) as the timeout threshold changes (x-axis). We choose a BGP inactivity timeout of 30 days (vertical line).

4.2 Establishing BGP lifetimes

We aggregate BGP data (§3.2) at daily granularity, consistently with the resolution available for administrative lifetimes. For each ASN, we consider the start of a *BGP lifetime* the first day we see it in BGP AS paths. However, differently from the administrative dimension, there is no reference concept to leverage to separate periods of BGP activity of an ASN into distinct lifespans. In addition, establishing the end of an ASN lifespan when such ASN is not seen in BGP for only 1 day would be misleading, since it is normal for a BGP speaker to temporarily stop originating prefixes or transiently disappear as a transit in preferred routes (*e.g.,* during an outage). Therefore, in order to introduce the concept of ASN "activity" in BGP for juxtaposition against the administrative dimension, we establish a timeout threshold.

We observe the distribution of per-ASN, activity time gaps with a daily granularity (Figure 3, red line) and select an arbitrary *activity timeout* threshold of 30 days, which is approximately where the "knee" of the CDF of activity time gaps starts and corresponds to 70.1% of the distribution. That is, we consider an ASN to start a new operational lifespan only if it reappears in BGP after > 30 days of inactivity. To further understand the implications of picking this threshold, we also look at the number of operational lives that a timeout value would cause to exist within the same administrative lifespan. We consider the "canonical" case for an administrative lifetime to contain at most 1 operational life and we thus compute the distribution of administrative lives that contain one or less operational lives (blue dotted line in Figure 3). Our 30 days threshold well fits the area where this CDF starts flattening and corresponds to 83% of the administrative lifetimes having only one or less operational lives. We obtain 152,926 BGP lifetimes for 96,391 ASNs, compared to 126,953 administrative lifetimes for 106,873 ASNs in the delegated files. In Appendix C we show the (minimal) impact on the rest of our analysis of varying this activity timeout.

```
# Administrative Dataset
{
    "ASN":205334,
    "regDate":"2017-09-20",
    "startdate":"2017-09-20",
    "enddate":"2021-02-11",
    "status":"allocated",
    "registry":"ripencc"
},
# Operational Dataset
{
    "ASN":205334,
    "startdate":"2017-10-05",
    "enddate":"2017-10-23"
}
```

Listing 1: Examples from our Administrative and Operational datasets. The snippets show the records for ASN 205334. The first one represents its administrative life: the AS has been registered and allocated by RIPE NCC in 2017-09-20 and deallocated on 2021-02-11. During that period, it was active in BGP from 2017-10-05 to 2017-10-23.

5 A BIRD'S EYE VIEW

In this section, we take a look at global and per-RIR trends. We present insights that emerge from a bird's eye view of the data, such as a large number of ASNs never used; in Appendix A we provide further insight into historical trends. In the next section (§6), we instead delve into an in-depth analysis.

A better understanding of regional trends. We find that by using our newly-built administrative and operational lifetime lenses we can better estimate trends (*e.g.,* compared to [31]). Figure 4 shows the count of alive ASNs per day, per RIR and overall: administrative and operational data are respectively depicted with solid and dashed lines; for the overall lines, we use the y-axis on the right side.[7] While all RIRs show a growing trend, RIPE NCC exhibits a much faster growth than the other RIRs since the very beginning of our observation period in 2004. At that time RIPE NCC had ten thousand less ASNs than ARIN, but in 2012 it surpassed ARIN, becoming the registry with the largest number of alive ASNs. Note that in public reports at [31] this overtaking is estimated to happen 4 years later, around 2016, since their methodology counts all ASNs ever allocated, including those that were later de-allocated (*i.e.,* returned to the pool of available resources or in transition (reserved) status). Moreover, when comparing the administrative and operational lives, the graph reveals that, in the operational perspective, RIPE NCC surpassed ARIN much earlier: in 2009 compared to 2012. In Appendix A, we show how this data, when broken down by country, provides insight into the expansion of Internet infrastructure in different countries and regions of the world over the years.

Many allocated ASNs are not operationally alive. The graph in Figure 4 also highlights that there is a significant gap between the two *overall* (BGP and administrative) lines, *i.e.,* many allocated

[7] Figure 13 in the Appendix shows the same data using a single axis.

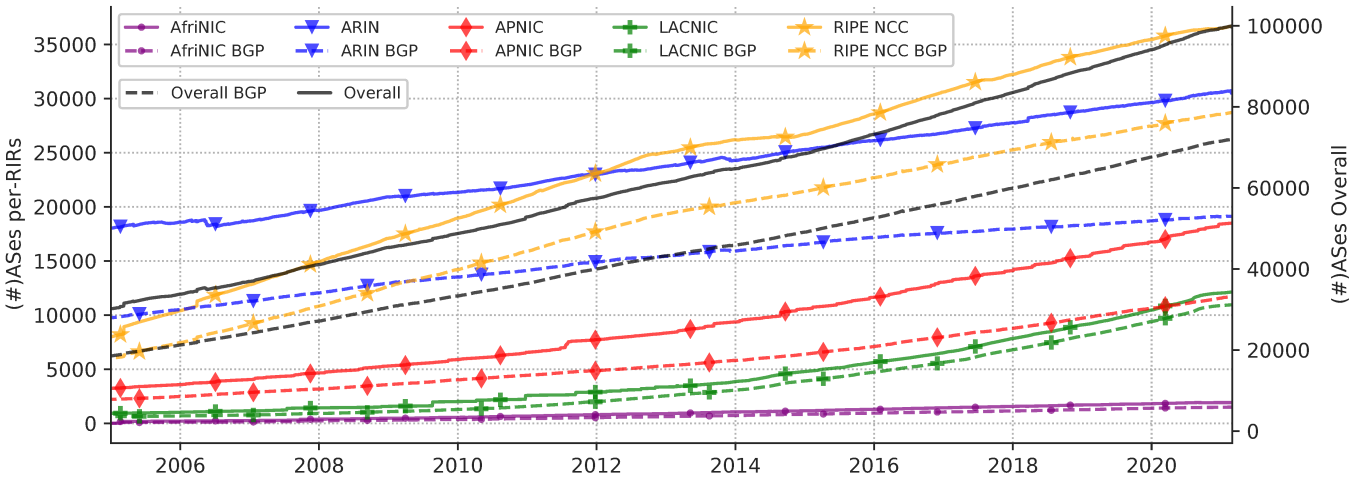

Figure 4: Administrative vs BGP lives: number of ASNs per day that are *administratively* (solid lines) and *operationally* (dashed lines) "alive", per RIR (colored) and overall (black). There is a significant and increasing gap between the number of ASNs in the two dimensions, with many ASNs allocated that are not alive in BGP. The growth of individual regions and the change in their proportions over time is visible (*e.g.,* RIPE NCC surpasses ARIN in terms of alive ASNs on BGP in 2009, and only in 2012 for allocations).

Table 2: Number of administrative and operational lifetimes per ASN.

	1 life		2 lives		>2 lives	
RIR	Adm.	Op.	Adm.	Op.	Adm.	Op.
AfriNIC	96.7%	78.6%	3%	12.5%	0.3%	8.9%
APNIC	93.2%	76.9%	6.1%	14.5%	0.7%	8.6%
ARIN	71.9%	65.8%	21.9%	22.4%	6.2%	11.8%
LACNIC	98.4%	88.4%	1.5%	7.9%	0.1%	3.7%
RIPE NCC	84.4%	76.2%	14%	15.0%	1.6%	8.8%
Total	84.1%	74.3%	13.4%	15.8%	2.5%	9.9%

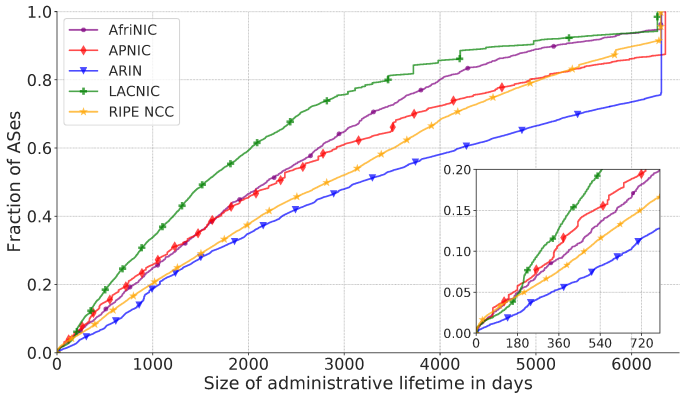

Figure 5: CDF of the duration of the administrative lifetimes per RIR. In the bottom right corner: zoom of the CDF focused on the fraction of ASes with shorter life (between 0 and 2 years).

ASNs are not used in BGP. In March 2021, this gap consisted of more than 27,800 ASNs, meaning that almost 28% of all allocated ASNs are not active in BGP (*i.e.,* . have not appeared in BGP announcements for at least 30 days). In §6.3 we analyze this phenomenon in detail and identify a set of causes.

RIRs still make ASN re-allocations. Most (84.1%) ASNs are never re-allocated. However, RIRs exhibit substantially different behaviors with respect to the reuse of ASNs: Table 2 ("Adm." columns) shows, for each RIR, how many ASNs have been allocated once, twice, or more. ARIN and RIPE NCC, re-allocate significantly more than the other RIRs, especially for ASNs that are re-allocated more than once: intuitively, being the two oldest and largest (by total ASNs) RIRs, there is a higher probability their ASNs are re-used. In addition, RIPE NCC and ARIN have more aggressive resource reuse policies [11], which can impact the reuse rate of those RIRs (see Appendix B for more details). However, as 32-bit ASNs became available in 2007—thus making AS numbers an extremely abundant resource—re-assigning previously used numbers would seem unnecessary and potentially at risk of creating conflicts with stale router configurations or routing policies that operators fail to update—a phenomenon we characterize in §6.2. Nevertheless, we observe this practice in all RIRs. A possible explanation is that 16-bit numbers are still a precious resource; we provide more insight about possible issues with 32-bit AS numbers in §6.3.

Many ASN allocations are short-lived. A large fraction of ASNs have a long life (CDF in Figure 5): more than *5 years* between 65% (ARIN) and 44% (LACNIC) and more than *10 years* between 42% (ARIN) and 19% (LACNIC). However, more interestingly, a significant portion of ASNs do not last more than 1 year! This fraction is higher in the 3 smaller RIRs (LACNIC 13%, APNIC 11%, AfriNIC 9%, versus RIPE NCC 8%, and ARIN 6%). However, when we break down the life duration by the birth year (Figure 14 in

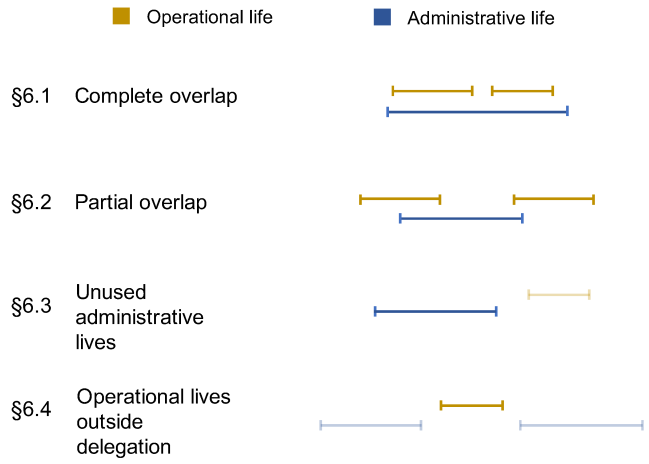

Figure 6: Taxonomy of behaviors that it is possible to observe when looking at individual ASNs through our compound lenses. The golden (blue) lines represent administrative (operational) lifetimes. The third and fourth cases show faded out lines representing lifetimes whose presence does not alter the specific case.

Appendix A shows a detailed sequence of boxplots), we find that, starting from around 2010, the life expectancy becomes similar across all RIRs, suggesting that in the last decade it has reached a certain stability in all RIRs. We also find that some short-lived ASNs are likely due to operational issues with 32-bit ASNs experienced by network operators (see §6.3 for more details). As RIRs started delegating 32-bits ASNs in 2010-2011, from then on they all have a significant share of ASNs with short administrative lifetimes.

The deployment of 32-bit ASNs is highly diverse across RIRs. Separating the allocations of 16- and 32-bit ASNs we can see how the registries managed the 16-bit ASN exhaustion and the transition to 32-bits. (Figure 12 in Appendix A shows per-day allocation status of 16- and 32-bit ASNs over time for each RIR). Unexpectedly—despite still being the 2nd largest RIR—ARIN is currently the fourth registry by 32-bit allocations and it only ramps up allocating these resources around 2014, several years after RIPE NCC, APNIC, and LACNIC. Still, in 2020, around 30% of ARIN's new allocations were 16-bit numbers—a completely different behavior compared to the younger registries (APNIC, LACNIC, AfriNIC) where 16-bit ASNs represented only between 1% and 1.7% of all the allocations each of them made in 2020. In Appendix A, we analyze the behaviors related to the 16-bit ASNs exhaustion in more detail.

6 JOINT ANALYSIS OF ADMINISTRATIVE AND OPERATIONAL LIVES

We now align the two lenses we have built in §4.1 and §4.2 in order to look at individual ASNs when bringing into focus both the administrative and the operational perspectives across time. Jointly looking at them provides an opportunity to better understand operational practices and identify anomalies. We first present a taxonomy of behaviors that it is possible to observe for each ASN when looked through our compound lenses. We then discuss representative examples and novel findings for each of these categories.

Table 3: Distribution of the 4 categories in our taxonomy illustrated in Figure 6.

Category	Adm. lives	Op. lives
§6.1 - Complete overlap	99,790	130,397
§6.2 - Partial overlap	4,434	5,434
§6.3 - Unused administrative lives	22,729	0
§6.4 - Op. lives outside delegation	0	2,382
Total	126,953	138,213

We classify ASNs into four different categories depending on how the administrative and operational lives compare, taking the administrative life as the primary reference. Figure 6 provides a graphical representation of the four categories and Table 3 shows the count and percentage of ASNs in them. The fourth category of ASNs that have an operational life in BGP without being allocated for the duration of that operational activity (*i.e.,* the operational life is outside any administrative life) may have a disjoint administrative life at another point in time that would fall in one of the 3 categories concerning administrative lives. The four categories in our taxonomy are the following:

(1) **Complete overlap:** This is the canonical case, where an operational lifetime happens entirely *within* the time that an ASN is in an allocated state. 78.6% (99,790) of the administrative lives fall in this category. However, we observe large variations *(i)* in the ratio between an operational lifespan and its corresponding administrative lifespan, and *(ii)* in the number of operational lifetimes within the same administrative lifetime. In §6.1 we dive into the range of behaviors that we observe in this category and the anomalies linked to malicious behavior that we find.
(2) **Partial overlap:** In this case, for a given ASN, we see an operational lifetime overlapping with an administrative lifetime but starting before and/or ending after it. 3.4% (4,434) of the administrative lives present this behavior. In most cases the operational life beginnings and end are close to the related administrative delegation indicating just a slow synchronization of the two dimensions. In §6.2 we describe more in detail our findings related to partial overlap.
(3) **Unused administrative lives:** These are administrative lifetimes with no BGP activity overlapping with them. Overall almost 18% (22,729) of administrative lives fall in this category. This behavior is partially explained by the limited visibility of ASNs in the BGP activity captured by the RouteViews and RIPE RIS collecting infrastructure, especially for the China region, the utilization of sibling ASNs, and issues in the deployment of 32-bit ASNs. We analyze and provide more detail on this category in §6.3.
(4) **Operational lives outside delegation:** We find a total of 1667 ASNs in this last category. In particular, we discover 799 ASNs that appear in BGP entirely outside of administrative lifetimes and 868 ASNs that are used in BGP for which there is no record of administrative delegation at all by any RIR in the entire 17-years period of examination. We find cases of malicious

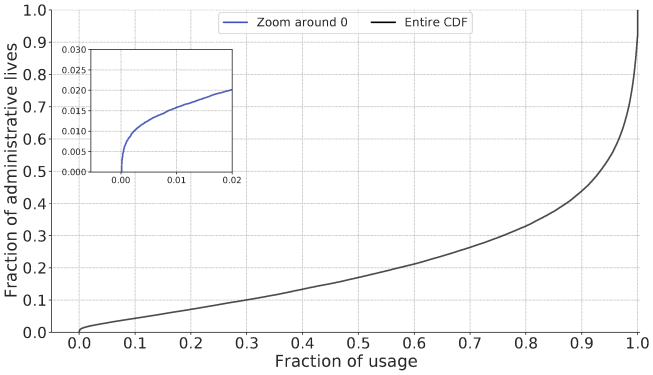

Figure 7: CDF of the usage of administrative lifetimes fully containing operational lifetimes, computed as the ratio between the sum the operational lifetimes an administrative lifetime contains and its duration.

behavior in the first category, and we identify some reasons for the second one. In §6.4 we describe each of these behaviors in detail.

6.1 Complete overlap

This is the most common case, accounting for 78.6% (99,790) of all the administrative lifetimes.

6.1.1 Lack of full utilization. Figure 7 shows the CDF of the utilization of each administrative life, computed as the ratio between the sum of the operational lifetimes an administrative lifetime contains and its duration. The majority of the administrative lives (70%) are heavily used (more than 75% of their duration) but a close to full usage happens in less than half of the cases (only 45% have a usage greater than 95%). On the contrary, many allocations are heavily under-utilized (*e.g.,* 10% are less than 30% utilized). We analyzed the causes of under-utilization, and found evidence of *(i) late deallocation*, *(ii) sporadic/intermittent use*, and *(iii) largely spaced operational lives*. Below we characterize and provide examples of each of these three behaviors.

Late deallocations. One of the main reasons for the lack of full operational utilization of delegated ASNs is the significant delay in the deallocation of ASNs when they are not operationally active. We find that it often takes months[8] for an ASN to be deallocated since its last day of BGP life: the median for APNIC ASNs is more than 6 months, and more than 10 for all the other RIRs, with AfriNIC's median value being almost a year and a half (530 days). This behavior highlights a potential security problem, which we discuss later, since these resources can be vulnerable to squatting attacks. Delays are also common, though less significant, in the start of operational activity in BGP after an ASN has been allocated: the median is greater than a month for all RIRs.

Sporadic/intermittent use. Another cause of lightly-used administrative lives is the intermittent behavior of BGP activity of some ASNs. The vast majority (84.1%) of the administrative lives that fully overlap with BGP activity actually contain only one operational life. Another significant fraction (10.4%) contains only two operational lives and—despite our 30-days threshold—5.4% has two or more lives. Surprisingly, 287 ASNs have more than 10 operational lives. We further investigate these ASNs and find that the majority of them (153 out of 287) have sibling ASNs, *i.e.,* they are part of an organization that manages multiple (sibling) ASes. This suggests that routing policies of large operators (*e.g.,* the same routes might be propagated using their siblings' ASNs, depending on internal routing adopted by the operator) are a possible explanation for sporadic BGP activity. In addition, we manually verify that other ASNs in this category are intermittent "by design": For example, AS37095 (African Network Operators' Group - AFNOG) and AS24555 (Asia Pacific Network Operators Group) are only used by the two network operator groups during their conferences or other events.

Largely spaced operational lives. A third reason causing under-utilization of ASN administrative lives are ASNs having very distant operational lives within the same ASN administrative allocation. Specifically, looking at administrative lives with more than one operational life, we see that 3,789 (23.9%) of them have operational lives more than 365 days apart. While this behavior might be due to organizational or operational changes within a company (*e.g.,* an AS going through changes of providers or in the arrangements with its provider—such as letting a provider announce its space in BGP on its behalf—we find several episodes of malicious activity within this behavior, which we discuss in the next paragraph.

6.1.2 Squatting of dormant ASNs. In *ASN squatting*, an attacker originates BGP announcements of prefixes using an ASN that it does not hold. The squatted ASN is either *(i) dormant*, *i.e.,* allocated but not used to advertise prefixes for long periods, or *(ii)* not allocated at all. This behavior is often associated with malicious purposes, such as announcing squatted prefixes[9] (*e.g.,* for spamming from non-blacklisted address blocks) or hijacking prefixes[10] (which enables various types of attacks). By originating from a different ASN than its own, the attacker tries to disguise their "BGP identity" [34]. For the same attacks, the attacker could also use its own ASN or one it *hijacked* from another organization that was allocated and active on BGP. However, using a dormant/unallocated ASN offers the advantage that potentially there is no owner to notice the event (similarly with property squatting).

We conjecture that, by leveraging the lens of combined administrative-operational lifetimes, squatting of dormant ASNs would result evident in extreme cases. The intuition, is that such attacks should happen after a long time of inactivity and for a short period of time compared to the whole administrative lifespan (*i.e.,* the operational life related to these squatting events will be very short compared to the administrative life of the ASN and far in time from the previous operational lifetime). To test our hypothesis, we set two parameters to detect possible malicious activity of dormant ASNs:

[8]We perform this analysis only on the administrative lives that end before the last day of our time frame of analysis, March 1, 2021.

[9]Prefixes advertised by a malicious actor that were allocated to other organizations that were not advertising them in BGP.

[10]Prefixes advertised by a malicious actor that were allocated to and are covered by BGP announcements of other organizations.

- A *period of inactivity* (while allocated) longer than 1000 days, either since the start of the administrative allocation or between operational lives.
- A *"relative duration" of the post-dormant operational life* (after being inactive in BGP for 1000 or more days and computed as its lifespan divided by the lifespan of the corresponding administrative lives) set to 5%.

Note that these thresholds are arbitrary *by design*, since here we are interested in simply testing our intuition through manual investigation. We find 3,051 operational lives matching our simple filter. We semi-automatically inspect them by counting the daily number of prefixes originated by BGP announcements of those ASNs, and checking their upstream to look for well known malicious actors. We successfully identify many suspicious cases, some of which we are able to cross-validate through external sources, finding at least 76 confirmed cases using information collected from network operators' mailing lists such as NANOG [53], Twitter alerts by network security groups such as Spamhaus [59], routing monitors such as BGPmon [4], and previous work [72]. Unfortunately, broad ground truth about hijacks is not available, thus we cannot quantify in detail how many of these cases are malicious. We confirm as many cases as possible using the sources cited above.

To illustrate this phenomenon, Figure 8 shows the number of prefixes originated over time by a subset of these ASNs, providing a visualization of the concept of the awakening of dormant ASNs (*i.e.,* not previously announcing prefixes and not seen in BGP for a long period of time). Furthermore, the figure shows that some hijacks happen simultaneously and we verify those prefix announcements share the same upstream provider (next hop in BGP), suggesting coordination of these attacks. For example, the second spike of AS10512 in the figure, represents a prefix hijacking event disclosed on the NANOG mailing list (the mailing list of North American operators) where one of the victims was Spectrum, a major broadband provider in the U.S. [52]. Even if AS10512 was allocated for more than 17 years (from 2003-11-20 to 2021-03-01), in BGP it was active for only 31 days, from 2017-12-08 to 2017-12-16 and from 2017-12-18 to 2018-01-09. Both periods match the spikes visible in Figure 8. In the second one, AS10512 suddenly originated 60 /16 prefixes for a short period, also causing (Sub)MOAS conflicts[11] for some of them, including prefixes originated by Spectrum (AS11426). In other words, AS10512 was squatted and used to perform BGP prefix hijacking attacks. The other ASNs in Figure 8 show similar behavior in terms of number of prefixes announced and in some cases also generate (Sub)MOAS events. We find that 2 of these ASNs are in the dataset of potential "serial" BGP hijackers created by Testart et al. in [72].

Some of the ASNs we pinpoint (including AS28071 and AS7449 in Figure 8), to the best of our knowledge, have not been previously identified as involved in these type of activities. Interestingly, we find that AS7449, which is unusually active in the same period AS10512 is, shares with it—in the BGP announcements of these events—the same direct upstream, AS203040, an ASN notoriously known as a "BGP Hijack Factory" [51]. It is thus most likely that AS203040 generated and shared with its neighbors forged BGP announcements with these (squatted) ASNs as origins and itself as the first hop, disguising itself as their transit. We identify a similar attack pattern for AS28071 and AS262916 (a well known spammer, reported in 2014 by BGPmon [12]), visible in Fig. 8 to be suddenly alive in BGP between 2013 and 2014: through inspection of the AS path in related BGP announcements, we learn they appear to share the same direct upstream—AS52302—during these activity spikes. Searching for this ASN, we find validation of its malicious behavior in the Latin America operators mailing list [46].

However, not all of these malicious events show a sudden increase in the number of prefixes originated per day, making it more challenging to detect them by solely studying their BGP activity without the allocation context. For example, between April and July 2020, 31 ASNs woke up almost simultaneously after several years of inactivity and started announcing each a few /20 prefixes that they never had announced before. We verified these announcements were also malicious, as they involved upstream ASNs known for this type of attacks [71].

Summing up on squatting of dormant ASNs. These case studies show that by using detection parameters that combine the administrative and operational perspectives it is relatively easy to put into focus malicious activity. Our newly-constructed lens could for example provide additional "classification features" for machine-learning based detection approaches. However, our study does not show to which extent and with which accuracy detection would be possible. As previous work on detecting BGP hijacking activity shows [72], it is hard to disambiguate legitimate operations exhibiting irregular/unusual behavior—explainable with traffic engineering, BGP blackholing, *etc.*—from malicious activity. Future work specifically focused on detection would need to rely on ground truth for all the events related to previously dormant ASNs, which is currently not available.

6.2 Partial Overlap

This category (second from the top in Figure 6), includes all administrative lives that have an operational life starting before and/or ending after it. They represent only 3.4% (4,434) of all administrative lives that we observe in 17 years of data. We find two benign reasons that explain most of the cases in this category and are described below.

Operators' dangling announcements. Most cases, (2,840, *i.e.,* 64% of all the administrative lives in this category) of partial overlap are due to operational lives continuing beyond the deallocation of their ASNs. The most probable explanation for these cases is the lack of reconfiguration of the routers (*e.g.,* by a provider of the AS). We study the size of ASes exhibiting this behavior using CAIDA ASRank historical snapshots [23] to retrieve their customer cone [48]—the set of ASes that can be reached from them following the customer links in their BGP paths. These ASNs are predominantly small: 95% of them have no customers. Thus, these dangling announcement likely come from manual router configurations that were not updated. Another possible cause of this behavior are stuck routes, where one of the ASNs in the path, does not record a withdraw update, therefore continuing seeing a path that should not exists anymore [17, 22]. While dangling announcements are a phenomenon known by registries, they constitute strong evidence

[11] Events in which two ASNs originate the same (MOAS) or overlapping (SubMOAS) prefixes.

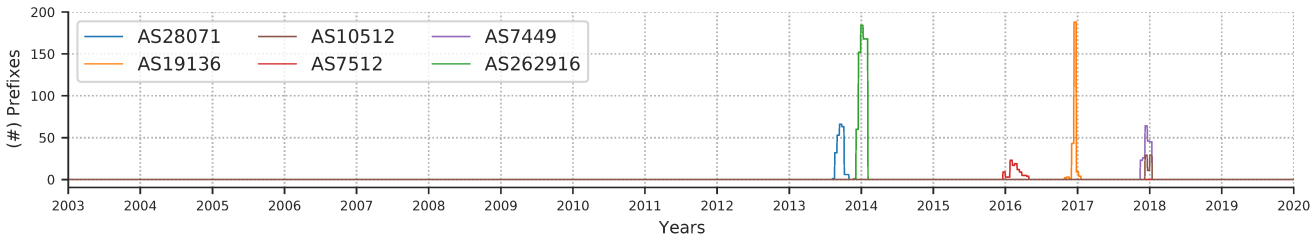

Figure 8: Number of prefixes originated by ASNs that suddenly "wake up" on BGP after years of inactivity (while staying allocated for the entire time). Our findings provide evidence of these events being related to malicious ASN squatting perpetrated in the context of BGP prefix hijacking attacks.

against re-use of ASNs. In our exchange with RIRs, we learned about cases where an RIR had to keep a deallocated ASNs in *reserved* status instead of putting it back in the available pool because of remaining BGP announcements with that ASN. An example is ASN 43268, which was allocated from 2007-07-05 to 2014-12-29 but appears in BGP announcements for almost 2 years after being deallocated (until 2016-09-01), prompting RIPE NCC to keep the ASN out of the available pool during that time.

Late allocations by RIRs. 1,594 ASNs start announcing prefixes in BGP before being allocated by an RIR. However, only 631 of them start announcing before the registration date shown in their respective allocation data. We find these mismatches only last a few days, suggesting their cause is due to a lack of synchronization between when RIRs communicate to the operator the assigned ASN and when they publish the allocation in their delegation files.[12] While this behavior seems of negligible importance, it has significant implications when hypothesizing to use delegation files as reference data for detecting potential misconfiguration and malicious behavior, which we discuss later in §9.

6.3 Allocated but unused administrative lives

No BGP activity is globally observed for a sizable fraction of administrative lifetimes. In total, for 22,729 (17.9%) administrative lives we do not find any BGP activity in our data during their lifespan. This phenomenon happens for 21,431 delegated ASNs, which is 20.7% of the total. Furthermore, 63% (13,407) of ASNs in this category have been allocated but are *never* seen in our BGP data in the entire 17-years period. We note that APNIC allocates entire blocks to National Internet Registries (NIRs), who perform individual allocations that we cannot track (*i.e.,* we consider all ASNs in the allocated block to have an administrative life). However, even if we do not count APNIC allocations, there are still 18,211 lives, allocated by the other 4 RIRs and never globally seen on BGP. This is surprising given that, according to RFC 1930, which provides the baseline guidelines RIRs follow for creating and delegating ASNs (see Appendix B for more details), *"an AS must be used for exchanging external routing information with other ASes"* [36].

[12]RIPE NCC stands out from other RIR by exhibiting an extremely large median value of 518 days between the start of ASN operation in BGP and the ASN appearance in delegation files. After manual investigation, we find that this is due to very old ASN resources (*i.e.,* from 1984-03-05 to 2002-09-06), which RIPE NCC added to its delegation files in bulk much later than the date appearing in their "registration date" field.

To characterize unused administrative lives, we start by inspecting their duration. Figure 9 shows the CDF of the duration of unused administrative lives by RIR. Interestingly, only a short portion of these lives are short-lived: depending on the RIR, only between 14.9% (ARIN) and 45% (LACNIC) of these ASNs had an administrative life lasting less than 1 year. We instead find that the majority of unused lives last multiple years, with a significant fraction being allocated for the entire observation period (the spikes at the end of each distribution)

Our further analysis of unused administrative lives suggests that *(i)* some of those ASNs might be used but are not globally observable in BGP, while others *(ii)* are actually unutilized for various reasons, including the use on the public Internet of sibling ASNs and the failed deployment of 32-bit ASNs. We discuss this analysis in the next paragraphs.

Disproportionate fraction of allocated-but-unobserved ASNs from China. China has a disproportionate fraction of its delegated ASNs that we do not observe in our BGP data. The BGP data collection infrastructure we use has varying levels of visibility depending on the topological and geographical location of ASes that share their BGP announcements with collectors. Nonetheless, we would expect only a small number of (likely transit) ASes impacted by limited visibility, but not such a large-scale phenomenon as the case with Chinese ASNs: Among the top-10 countries by number of unused administrative lives, China is by far the country with the largest fraction of its administrative lives being "allocated-but-unobserved", with 50.6% of all allocated ASNs being unobserved in BGP during the allocation lifetime compared to values below 15% for the runner up countries. Moreover, Chinese allocated-but-unobserved administrative lives represent more than 27% of all the allocated-but-unobserved lives in the APNIC region, even if China has only 10% of APNIC ASN allocations. The other top-10 countries exhibit a much smaller contrast. The next largest is France (14.5% of allocated-but-unobserved), holding—of all administrative lives in the RIPE NNC region—7.9% of allocated-but-unobserved lives but only 4.85% of the allocated (either observed or unobserved). Most other countries have comparable shares of allocated-but-unobserved and all delegations in their respective region. However, Russia stands out for the opposite reason, with a far smaller percentage of allocated-but-unobserved (8.12%) administrative lives compared to all allocated ones (16%), respectively in the RIPE NCC region. We conjecture that the large fraction of unused

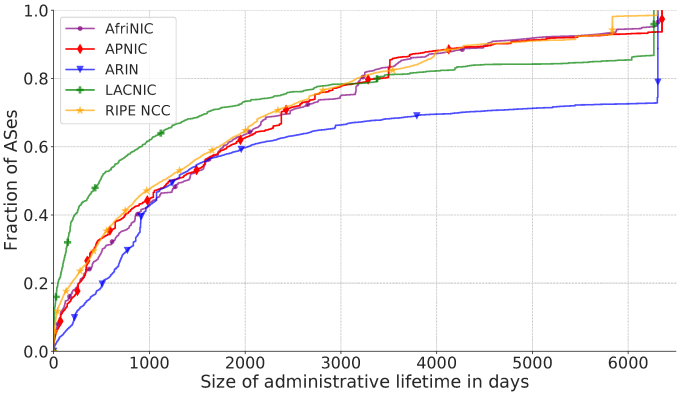

Figure 9: Distribution of lifetime for the never used ASNs.

ASNs from China is due to how routing is managed in the country: it is possible that several ASNs within the Chinese national AS-level topology are stripped from the AS-paths (*e.g.,* through route aggregation) by their upstream providers before being propagated to the rest of the Internet (where the RouteViews and RIS vantage points are located).

Unused ASN with sibling ASNs in use. Several organizations appear to keep their ASN allocations (and paying the negligible fee) even if they do not use an ASN in BGP—thus either not using it at all or using it only internally. We observe that a large fraction of allocated-but-unobserved ASNs have sibling ASNs, that is, the organization owning them owns also other ASNs. Organizations presenting such behavior include government organizations, such as the US Department of Defense and Air Force—for which we observe only around 40% and 45% of their allocated ASNs respectively—and companies that received large blocks of ASN allocations in the early years, such as Verisign and France Telecom (currently Orange)—which use only 24% and 20% of allocated ASNs respectively.

Challenging deployments of 32-bit ASNs. We examine short-lived unused administrative lives and find that the vast majority of them are 32-bit ASN allocations. Among the unused administrative lives shorter than a month (31 days), 32-bit ASNs represent 92.6% for APNIC, 81% for AfriNIC, 87.3% for RIPE NCC, 65.2% for ARIN, and 38% for LACNIC. By leveraging ARIN's WhoWas service [9], which provides historical information about expired allocations made by ARIN, we investigate if these short-lived allocations are linked to operational issues: We check which organizations were responsible for a random half of the 101 ARIN short lifespans. We then search for the organization names in the list of currently allocated ASNs, and we find that 86% of these organizations have been assigned 16-bit ASNs right after the end of the previous (short-lived) 32-bit ASNs allocation. This finding suggests that short administrative lives that we do not observed in BGP might potentially be caused by operational issues with the deployment of 32-bit ASNs.

6.4 Operational lives without allocation

We identify 1,667 ASNs announcing in BGP without an overlapping administrative lifetime. Within this category, we find more evidence of abuse of unused resources (similar to §6.1.2) and ample evidence of misconfigurations. We split them in two sub-categories: 799 ASNs that at a certain point in time were allocated but had at least one BGP life entirely outside of any administrative life and 868 ASNs that have *never* been allocated. Note that we exclude from our analysis "bogon" ASNs normally filtered by operators, *i.e.,* ASNs reserved for special use [1, 29, 40, 44, 50, 75].

More BGP hijacking. Examining ASNs in the first sub-category, which are used in BGP outside their administrative allocation (*i.e.,* after being deallocated), we identify 9 prefix hijacking events where these ASNs were used as origins. We were able to corroborate these events through the same data sources mentioned in §6.1.2. Interestingly, we find that these events are not necessarily far from the closest administrative life but they are always far in time from the last (if ever) seen BGP life. E.g., we see AS12391 originating two /16 blocks and a /18 block (with AS197426 (Bitcanal) as upstream) 3 days after the deallocation of its ASN but 3,898 days after its previous operational life. Note that, differently from the cases we discover and highlight in §6.1.2, these ASNs were not allocated at the moment they were abused. This means that checking the status of these resources on the delegation files could have helped in identifying and preventing these squatting events.

"Fat-finger" misconfigurations that last months. When investigating the 868 ASNs that show BGP activity despite *never* being allocated in our entire 17-years observation period, we identify significant instances of misconfiguration events. Of the ASNs never allocated that appear in BGP, only 427 are active for more than 1 day, 186 more than 1 month, and 15 more than 1 year. We manually investigate more than half of these ASNs and find 258 (29.7%) evident cases of misconfiguration. 76% of these misconfigurations involve an origin ASN similar to an ASN in the AS Path of BGP announcements usually the first hop (*i.e.,* the ASN after the origin): these errors are typically caused by a failed attempt of AS path prepending [16]. For example, in 42 cases we find in the AS path an ASN that is an exact repetition of the origin ASN, such as AS3202632026, where the first hop is AS32026. In the remaining 24% cases, we observe Multiple Origin AS (MOAS) conflicts involving ASNs that differ by 1 digit. Surprisingly these events can last several months. For example, AS419333 appears in BGP for almost 10 months (between Nov, 2017 and Sep, 2018) causing a MOAS with AS41933, IPRAGAZ-AS. Another example is AS363690 causing a MOAS with AS393690 for almost 7 months (between Nov, 2018 and Jun, 2019).[13]

Unallocated ASNs used internally leak to the global Internet. Among the "never allocated" ASNs, we also observed (unallocated) ASNs with very large numbers. We found that 472 (54.4% of the 868 never allocated) have more digits than the highest allocated ASN, which is 6 digits long. The majority of the events we could manually investigate appear to be the unintended consequence of benign behavior and often last months, if not years. For example, AS290012147 announced a /24 prefix for more than 2 years (between 2015 and 2017), which is covered by a /12 announced by AS701, held by Verizon. We collect all AS paths from BGP announcements including that ASN for a day (while it was announced) and find that they all have the ASN triplet {AS290012147, AS7046, AS701}. Since

[13] Note that an attacker might be able to carefully choose an ASN to squat that looks like a mistyped ASN of the victim. In the cases we investigated, we verified that the upstream ASNs in the AS paths match the upstreams of the corresponding legitimate ASN (*i.e.,* strongly suggesting that these are actual fat-finger mistakes).

both AS701 and AS7046 are held by Verizon, and AS701 announces the covering /12 prefix, it is very likely that such announcements are due to a misconfiguration "leaking" routes used internally by Verizon. Similarly, we find events associated with other large unallocated ASNs (such as AS499981773, AS3489671207, and AS12845938). Note that these are not "bogon" ASNs defined in RFCs for internal use but are actual valid ASNs that RIRs might allocate.

7 RELATED WORK

The allocation of Internet resources has been studied for a long time, however the focus has been on IP block allocations. Huston [26, 28, 32] has produced information on the total number of allocations of IPs along with per RIR allocation analysis: How many resources are allocated in the delegated files and how many of them are routed. With this analysis, Huston shows the increased rate of IPs allocation and gives insights on IPv4 address exhaustion. In [61], Richter et al. study IPv4 addresses exhaustion and how the evolution and management ecosystem created diverse realities in different regions. In [62], Richter et al. analyze the operational use of IPv4 addresses from the point of view of a large CDN and characterize behaviors revealing under-utilization in some regions and complete utilization in others. Starting from the delegated files, Meng et al. analyze the correlation between the allocation of IP blocks and their usage in BGP, discovering that most of the prefixes allocated between 1997 and 2004 appear as routed after 75 days and that 8% have not been used at all [49]. Sriraman et al. [70] analyze the fragmentation of the IP address space contrasting allocated blocks with block routed on BGP for a period of five years, finding that almost 90% of ASes with a provider-customer relationship do not share an address delegation relationship. Similarly, Heidemann et al. [37] use allocation data of IPs to assess that only 3.6% of these addresses are actually visible hosts. More recently, Dainotti et al. [18] proposed a taxonomy and a new method combining active and passive measurements to understand address utilization. They discovered that only 37% of the total number of IPv4 usable addresses are actually used, and that most of the unused blocks are in the US. Other work focuses on the effectiveness of bogon lists and on how to improve their use [10, 19, 21]. In particular, the most common problem is that these lists are usually not updated as soon as new allocations are made, and therefore valid routes can be filtered out. Vaidyanathan et al. [74] introduced in the bogon lists the semi-dark space, addresses that are not in operational use. All these works focus on IP allocation rather than ASes.

Concerning ASes, many works have studied specific aspects of AS behavior in BGP without considering ASN delegations and their administrative lives. Chang et al. [15] built AS-TRUST, a scheme to quantify the reputation of an AS based on BGP updates, showing that it is possible to improve BGP operations. Konte et al. build ASwatch, a system to find bulletproof hosting ASes based on network and connectivity features of ASes inferred from BGP data [43]. Since these works do not take into account ASN delegations, they do not evaluate AS behavior depending on allocation status, which would allow to discern ASNs that were previously delegated to another organization. In [72], Testart et al. build a supervised machine learning system to find ASes that persistently hijack BGP prefixes. In our work, we provide evidence that using both the administrative and the operational dimensions, it is possible to separate behaviors from different allocations (*i.e.,* different administrative lives of the same ASN), thus possibly better characterizing the overall AS behavior. We believe this approach can improve detection methods solely based on the operational activity. Huston [27, 30] has published analyses on ASN consumption and aggregated allocation. Other works on ASes analyze their connectivity structures [56, 67, 69]. In summary, most of the works on ASes are based on BGP data and their interconnections rather than the life of these resource allocations in the Internet and their effective use in BGP.

In 2005, Wilhelm and Uijterwaal correlated ASN delegations and their activity in BGP [77], However, in 2005 AfriNIC was just born and we are now able to analyze 17 years of data. Policies changed and extended delegated files carrying more information have been introduced, allowing us to better characterize what invalid resources are being advertised. Moreover, we introduce new concepts such as ASN delegated life, ASN BGP life and ASN usage and perform a longitudinal analysis on the correlation between administrative and BGP lives.

8 LIMITATIONS

ASN-level granularity. In our study, we work with ASN-level data. We do not look at the individual prefixes advertised by ASNs, except in few manual analyses to better understand and characterize our findings (as in §6.1.2 on ASN squatting). However, information about the announced prefixes may help to further build and characterize BGP lifetimes, *e.g.,* identifying different BGP lifetimes of the same ASN based on different sets of announced prefixes. E.g., in §4.2 we pick an arbitrary 30-days inactivity threshold to separate two operational lives. Using prefixes, we could consider both the inactivity period and the prefixes announced by the ASN to decide whether to start a new operational lifespan or not.

Visibility limitations. We can only infer the use of an ASN in BGP if the BGP announcements from that ASN reach a peer of the collecting infrastructure we use. The existing collecting infrastructures have several vantage points, but they are not uniformly distributed around the globe. Indeed there are jurisdictions such as China, that heavily control the local interconnection with the global Internet and where such measurement infrastructure is not present. This is a factor that can limit the inference of operational activities of ASNs in some specific geographical areas.

Collectors There are other BGP data collection infrastructures available, such as *e.g.,* from the Packet Clearing House project (PCH) [39]. However, adding further collectors is unlikely to significantly alter our findings, since—differently from BGP prefixes, which might not propagate far in the topology, or might be shared in private peerings, or might end up aggregated—the operational information we are interested in (AS numbers from BGP announcements) does propagate in the topology. An exception would be if *e.g.,* PCH or another BGP collecting infrastructure had a presence in China, where (see previous paragraph) we find a limitation due to likely a filtering of AS numbers; in that case, we might be able to observe Chinese ASNs that are never propagated to the rest of the Internet. We are not aware of public BGP data collection infrastructure with such coverage.

Private peering. Another issue we might encounter is ASNs not

visible in BGP because used for private peering. However, in the majority of such cases, we would expect the owning organizations to also use a second ASN publicly. If this was a significant phenomenon, we would find many unobserved ASNs to have siblings. In §6.3 we show that sibling ASNs are not significant in number and are not enough to explain the extremely large number of unseen ASNs we find.

9 DISCUSSION

In this paper we align two dimensions along which ASNs are visible across time: their administrative allocation by registries and their operational use in BGP. ASNs are a key Internet infrastructural resource and this link is crucial for the operation and security of inter-domain routing but has received little attention in the research community. The combination of the administrative and operational lenses that we build through our datasets allows us to characterize the different behaviors that stem from the *interaction* between ASN delegation and BGP, the policies set by Internet Registries, misconfigurations, and malicious behavior.

Contrasting the administrative and operational dimensions of an ASN, we find that even though most organizations receive an ASN allocation and then start operating in BGP, there is a large breadth of different behaviors. At the two extremes, we find ASNs that are delegated (for many years) that never appear in BGP, and ASNs that operate in BGP without being allocated at that time. In between we have BGP operation fully or partially covering the ASN allocation. These behaviors are shaped by 3 distinct aspects:

- **RIRs policies and management of ASN delegations:** Whether RIRs delegate in block or mainly single ASNs, the internal delegation process (and when ASN are included in delegation files), the reuse policies and re-allocation process of previously allocated ASNs, and the choice of delegating 16-bits vs. 32-bit ASNs, they all impact the usage of allocated ASNs in BGP. Therefore, further study of our dataset can help elucidate best practices for both the delegation and use of ASN resources and the broader impact of these policies in the Internet infrastructure and ecosystem.
- **Misconfigurations and mistakes in operational setting and in RIRs delegation process:** Many operational and administrative errors quickly show up as anomalous behavior when combining these lenses. Indeed we find that fat-finger errors are the largest contributor of ASNs seen in BGP that have never ever been allocated to an organization. When these fat-finger errors and other misconfigurations relate to the origin AS, access to authoritative records of the correct ASN as origin of a given prefix would allow to verify the information in BGP and limit the spread of invalid announcements. Thus, if ASes have properly issued Route Origin Authorizations (ROAs) in the Resource Public Key Infrastructure (RPKI) for their prefixes, the spread of errors and misconfigurations would be limited when networks in the path drop RPKI-invalid announcements, *i.e.,* implement RPKI filtering.
- **Malicious behavior:** By studying the usage of ASNs in BGP during and after administrative allocation we are able to spot many indications of malicious behavior. There is much further work to do to characterize all the malicious behavior that is detected with these combined lenses. However, as a high-level conclusion from our manual analysis, hijackers are ahead of us: they carefully pick dormant or previously allocated ASNs to make their attacks stealthier (*i.e.,* mostly avoiding picking never-allocated ASNs, which we instead see in misconfigurations). Similarly to the case of misconfigurations though, when unallocated ASes are used as origin, if the victims of attacks had properly registered ROAs providing an authoritative record of the ASN authorized to announce as origin a given prefix, networks dropping RPKI-invalid would limit the spread of this type of attacks.

Practical relevance: We argue that this dual-lens has operational value to reduce the spread of misconfigurations in BGP (*e.g.,* by filtering all ASNs that are not delegated) and make malicious behavior, as well as operational problems (*e.g.,* the challenge with 32-bit ASNs), more visible. However, our study also highlights inconsistencies and behaviors—*e.g.,* mistakes and delays in the delegation files, dangling announcements after deallocation, large AS numbers "illegitimately" used internally and sometimes leaking—that should be addressed through policy and best practices in order to make delegation information more useful for operational purposes.

We make our polished datasets publicly available to the community for both reproducibility and for other works to leverage the administrative and operational lifetimes of ASNs in the Internet. We also intend to continue updating and publishing our datasets on a daily basis in order to facilitate near-realtime analysis and discussion around their potential for operational use.

Future work: In the future, we plan to use our dataset to further characterize aspects of ASN behavior that can extend our understanding of how the different organizations that make up the public Internet operate. As we explain in §4.1, APNIC also allocates blocks of ASNs to the NIRs, introducing uncertainty on when the ASNs are given to organizations. We plan to contact the NIRs and ask if it is possible to access their allocation information in order to keep track of when these resources are assigned to an organization. In §6.3 we suggest that the high number of unused AS numbers of China may be caused by routing management in the country; to better understand this phenomenon, one solution might be to survey IXPs and operators to understand their routing policies.

With respect to malicious behavior, in this study we do not aim to accurately detect attacks, but rather show the potential of our new "compound lens". In the future, we plan to focus on the development of a detection methodology that can take full advantage of our data to discover malicious events and misconfigurations.

Finally, we expect to extend our dataset to integrate other information shaping ASNs behavior: (1) information about sibling organizations in order to prune our correlation between administrative lifetime and BGP lifetime; (2) data from IP address delegations with the purpose of better characterizing the administrative dimension of a network; and (3) distinguishing between origination and transit BGP activity of an ASN to differentiate the role(s) an ASN has at different times of its BGP lifetime.

ACKNOWLEDGMENTS

We thank the anonymous reviewers for their thoughtful feedback. This work was partially supported by the MIT Internet Policy Research Initiative, William and Flora Hewlett Foundation grant 2014-1601. We acknowledge funding support from NSF Grant CNS

1705024. This work was supported in part by the MIUR under grant "Dipartimenti di eccellenza 2018-2022" of the Department of Computer Science of Sapienza University.

REFERENCES

[1] J. Abley and W. Sotomayor. May 2015. *RFC7534: AS112 Nameserver Operations.* https://tools.ietf.org/html/rfc7534

[2] AfriNIC. [n. d.]. *AfriNIC ftp.* Retrieved 2020-05-17 from ftp://ftp.afrinic.net/pub/stats/afrinic/

[3] AfriNIC. Sep. 2015. *afrinic-10years.* Retrieved 2020-05-11 from https://afrinic.net/ast/pdf/afrinic-10years-ab-sept-2015.pdf

[4] Cisco and/or its affiliates. 2020. *BGPmon.* https://bgpmon.net/

[5] APNIC. [n. d.]. *APNIC ftp.* Retrieved 2020-05-17 from https://ftp.apnic.net/stats/apnic/

[6] APNIC. 2020. *apnic erx resources.* https://www.apnic.net/manage-ip/manage-historical-resources/erx-project/erx-asn-transfer/

[7] ARIN. [n. d.]. *ARIN ftp.* Retrieved 2020-05-17 from ftp://ftp.arin.net/pub/stats/arin/

[8] ARIN. 2020. *erx resources.* ftp://ftp.arin.net/erx/asn/erx-asns.txt

[9] ARIN. 2021. *WhoWas Service.* https://www.arin.net/reference/research/whowas/

[10] Jon Arnold, Olaf Maennel, Ashley Flavel, Jeremy McMahon, and Matthew Roughan. 2008. Quantitative analysis of incorrectly-configured bogon-filter detection. In *2008 Australasian Telecommunication Networks and Applications Conference*. IEEE, 10–15.

[11] Number Resource Organization (ASO). 2010. *RIR Comparative Policy Overview 2010-03.* https://www.nro.net/rir-comparative-policy-overview-2010-03/#1-3-3

[12] bgpmon. Sep. 2014. *Using BGP data to find Spammers.* https://bgpmon.net/using-bgp-data-to-find-spammers/

[13] V. Cerf. Aug. 1990. *RFC 1174: IAB Recommended Policy on Distributing Internet Identifier Assignment and IAB Recommended Policy Change to Internet "Connected" Status.* https://tools.ietf.org/html/rfc1174

[14] Vint Cerf. Oct. 1969. *RFC 20: ASCII format for Network Interchange.* https://tools.ietf.org/html/rfc20

[15] Jian Chang, Krishna K Venkatasubramanian, Andrew G West, Sampath Kannan, Boon Thau Loo, Oleg Sokolsky, and Insup Lee. 2011. AS-TRUST: A trust quantification scheme for autonomous systems in BGP. In *International Conference on Trust and Trustworthy Computing*. Springer, 262–276.

[16] Shinyoung Cho, Romain Fontugne, Kenjiro Cho, Alberto Dainotti, and Phillipa Gill. 2019. BGP hijacking classification. In *2019 Network Traffic Measurement and Analysis Conference (TMA)*. IEEE, 25–32.

[17] Ben Cox. 2021. *Hunting down the stuck BGP routes.* Retrieved 2021-05-25 from https://blog.benjojo.co.uk/post/bgp-stuck-routes-tcp-zero-window

[18] Alberto Dainotti, Karyn Benson, Alistair King, Bradley Huffaker, Eduard Glatz, Xenofontas Dimitropoulos, Philipp Richter, Alessandro Finamore, and Alex C Snoeren. 2016. Lost in space: improving inference of IPv4 address space utilization. *IEEE Journal on Selected Areas in Communications* 34, 6 (2016), 1862–1876.

[19] D Dietrich. 2005. Bogons and bogon filtering. In *33rd meeting of the North American Network Operator's Group (NANOG 33)*.

[20] Christoph Dietzel, Anja Feldmann, and Thomas King. 2016. Blackholing at ixps: On the effectiveness of ddos mitigation in the wild. In *International Conference on Passive and Active Network Measurement*. Springer, 319–332.

[21] Nick Feamster, Jaeyeon Jung, and Hari Balakrishnan. 2005. An empirical study of" bogon" route advertisements. *ACM SIGCOMM Computer Communication Review* 35, 1 (2005), 63–70.

[22] Romain Fontugne, Esteban Bautista, Colin Petrie, Yutaro Nomura, Patrice Abry, Paulo Gonçalves, Kensuke Fukuda, and Emile Aben. 2019. BGP zombies: An analysis of beacons stuck routes. In *International Conference on Passive and Active Network Measurement*. Springer, 197–209.

[23] Center for Applied Internet Data Analysis based at the University of California's San Diego Supercomputer Center. 2021. *AS Relationships.* https://www.caida.org/catalog/datasets/as-relationships/

[24] American Registry for Internet Numbers. 2004. *New Statistics Format Available.* Retrieved 2020-05-05 from https://www.arin.net/vault/announcements/2004/20040108.html

[25] American Registry for Internet Numbers. 2009. *Extended Allocation and Assignment Report for RIRs.* Retrieved 2020-05-05 from https://www.arin.net/reference/research/statistics/nro_extended_stats_format.pdf

[26] Huston Geoff. [n. d.]. *Ipv4 stats.* Retrieved 2020-06-29 from https://www.potaroo.net/tools/ipv4/index.html

[27] Huston Geoff. Aug. 2005. *ASN stats.* Retrieved 2020-06-29 from https://www.potaroo.net/ispcol/2005-08/as.pdf

[28] Huston Geoff. Dec. 2003. *IPv4 - How long do we have?* Retrieved 2020-06-29 from https://www.potaroo.net/papers/ipj/2003-v6-n4-ipv4/ipv4.html

[29] Huston Geoff. Dec. 2008. *RFC 5398:Autonomous System (AS) Number Reservation for Documentation Use.* https://tools.ietf.org/html/rfc5398

[30] Huston Geoff. Jul. 2003. *ASN IPs stats.* Retrieved 2020-06-29 from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.108.5361&rep=rep1&type=pdf

[31] Huston Geoff. May. 2021. *The 32-bit AS Number Report.* Retrieved 2021-05-22 from https://www.potaroo.net/tools/asn32/

[32] Huston Geoff. Oct. 2008. *Confronting IPv4 Address Exhaustion.* Retrieved 2020-06-29 from https://www.potaroo.net/ispcol/2008-10/v4depletion.pdf

[33] Dan Goodin. Apr. 2018. *Suspicious event hijacks Amazon traffic for 2 hours, steals cryptocurrency.* Retrieved 2020-06-29 from https://arstechnica.com/information-technology/2018/04/suspicious-event-hijacks-amazon-traffic-for-2-hours-steals-cryptocurrency/

[34] Google. 2018. *The Hunt for 3ve Taking down a major ad fraud operation through industry collaboration.* https://services.google.com/fh/files/blogs/3ve_google_whiteops_whitepaper_final_nov_2018.pdf

[35] J. Haas and J. Mitchell. Jul. 2014. *RFC 7300: Reservation of Last Autonomous System (AS) Numbers.* https://tools.ietf.org/html/rfc7300

[36] J. Hawkinson and T. Bates. Mar. 1996. *RFC 1930: Guidelines for creation, selection, and registration of an Autonomous System (AS).* https://tools.ietf.org/html/rfc1930

[37] John Heidemann, Yuri Pradkin, Ramesh Govindan, Christos Papadopoulos, Genevieve Bartlett, and Joseph Bannister. 2008. Census and survey of the visible internet. In *Proceedings of the 8th ACM SIGCOMM conference on Internet measurement*. 169–182.

[38] Urs Hengartner, Sue Moon, Richard Mortier, and Christophe Diot. 2002. Detection and analysis of routing loops in packet traces. In *Proceedings of the 2nd ACM SIGCOMM Workshop on Internet measurment*. 107–112.

[39] Packet Clearing House. 2021. *Packet Clearing House.* https://www.pch.net/

[40] IANA. Aug. 2015. *Special-Purpose Autonomous System (AS) Numbers Created.* Retrieved 2020-05-17 from https://www.iana.org/assignments/iana-as-numbers-special-registry/iana-as-numbers-special-registry.xhtml

[41] Daniel Karrenberg, Gerard Ross, Paul Wilson, and Leslie Nobile. 2001. Development of the Regional Internet Registry System. *The Internet Protocol Journal* 4, 4 (Dec. 2001), 17–29. https://www.nro.net/development-of-the-regional-internet-registry-system/

[42] S. Kirkpatrick, M. Stahl, and M. Recker. Jul. 1990. *RFC 1166: Internet numbers.* https://tools.ietf.org/html/rfc1166

[43] Maria Konte, Roberto Perdisci, and Nick Feamster. 2015. ASwatch: An AS Reputation System to Expose Bulletproof Hosting ASes. In *Proceedings of the 2015 ACM Conference on Special Interest Group on Data Communication - SIGCOMM '15*. ACM Press, London, United Kingdom, 625–638. https://doi.org/10.1145/2785956.2787494

[44] W. Kumari, R. Bush, H. Schiller, and K. Patel. Aug. 2015. *RFC7607:Codification of AS 0 Processing.* https://tools.ietf.org/html/rfc7607

[45] LACNIC. [n. d.]. *LACNIC ftp.* Retrieved 2020-05-17 from https://ftp.lacnic.net/pub/stats/lacnic/

[46] lacnog. Oct. 2013. *secuestro de ruta.* Retrieved 2020-06-29 from https://mail.lacnic.net/pipermail/lacnog/2013-October/002622.html

[47] K. Lougheed and Y. Rekhter. Jun. 1989. *RFC 1105: A Border Gateway Protocol (BGP).* https://tools.ietf.org/html/rfc1105

[48] Matthew Luckie, Bradley Huffaker, Amogh Dhamdhere, Vasileios Giotsas, and KC Claffy. 2013. AS relationships, customer cones, and validation. In *Proceedings of the 2013 conference on Internet measurement conference*. 243–256.

[49] Xiaoqiao Meng, Zhiguo Xu, Beichuan Zhang, Geoff Huston, Songwu Lu, and Lixia Zhang. 2005. IPv4 address allocation and the BGP routing table evolution. *ACM SIGCOMM Computer Communication Review* 35, 1 (2005), 71–80.

[50] J. Mitchell. Jul. 2013. *RFC6996: Autonomous System (AS) Reservation for Private Use.* https://tools.ietf.org/html/rfc6996

[51] Nanog. Aug. 2017. *Hijack Factories: AS203418, AS205944, and AS203040.* Retrieved 2020-06-29 from https://mailman.nanog.org/pipermail/nanog/2017-August/191858.html

[52] Nanog. Jan. 2018. *Spectrum prefix hijacks.* Retrieved 2020-06-29 from https://mailman.nanog.org/pipermail/nanog/2018-January/193573.html

[53] Inc. NANOG. 2021. *NANOG.* https://www.nanog.org/resources/nanog-mailing-list/nanog-mailing-lists/

[54] RIPE NCC. 2021. *Routing Information Service (RIS).* Retrieved 2021-05-25 from https://www.ripe.net/analyse/internet-measurements/routing-information-service-ris

[55] Number Resource Organization (NRO). 2021. *RIR Comparative Policy Overview.* https://www.nro.net/policy/regional/rir-comparative-policy-overview/

[56] Ricardo Oliveira, Dan Pei, Walter Willinger, Beichuan Zhang, and Lixia Zhang. 2009. The (in) completeness of the observed internet AS-level structure. *IEEE/ACM Transactions on Networking* 18, 1 (2009), 109–122.

[57] Chiara Orsini, Alistair King, Danilo Giordano, Vasileios Giotsas, and Alberto Dainotti. 2016. BGPStream: a software framework for live and historical BGP data analysis. In *Proceedings of the 2016 Internet Measurement Conference*. 429–444.

[58] J. Postel and J. Vernon. Jan. 1983. *RFC 820: Assigned Numbers.* https://tools.ietf.org/html/rfc820

[59] The Spamhaus Project. 2021. *Spamhaus.* https://www.spamhaus.org/

[60] Y. Rekhter, T. Li, and S. Hares. January 2006. *RFC 4271:A Border Gateway Protocol 4 (BGP-4).* https://www.rfc-editor.org/info/rfc4271

[61] Philipp Richter, Mark Allman, Randy Bush, and Vern Paxson. 2015. A Primer on IPv4 Scarcity. *ACM SIGCOMM Computer Communication Review* 45, 2 (April 2015), 21–31. https://doi.org/10.1145/2766330.2766335

[62] Philipp Richter, Georgios Smaragdakis, David Plonka, and Arthur Berger. 2016. Beyond Counting: New Perspectives on the Active IPv4 Address Space. In *Proceedings of the 2016 Internet Measurement Conference*. ACM, Santa Monica California USA, 135–149. https://doi.org/10.1145/2987443.2987473

[63] RIPE. [n. d.]. *RIPE ftp*. Retrieved 2020-05-17 from https://ftp.ripe.net/pub/stats/ripencc/

[64] RIPE. 2020. *legacy-resources*. https://www.ripe.net/manage-ips-and-asns/legacy-resources/erx

[65] RIPE. 2020. *ripe erx resources*. https://www.ripe.net/manage-ips-and-asns/legacy-resources/erx/erx-transfer-of-as-number-registrations

[66] Eric C. Rosen. Oct. 1982. *RFC 827: Exterior Gateway Protocol (EGP)*. https://tools.ietf.org/html/rfc827

[67] Matthew Roughan, Walter Willinger, Olaf Maennel, Debbie Perouli, and Randy Bush. 2011. 10 lessons from 10 years of measuring and modeling the internet's autonomous systems. *IEEE Journal on Selected Areas in Communications* 29, 9 (2011), 1810–1821.

[68] RouteViews. 2021. *RouteViews Routing Table Archive*. Retrieved 2021-05-22 from http://www.routeviews.org

[69] Georgos Siganos and Michalis Faloutsos. 2004. Analyzing BGP policies: Methodology and tool. In *IEEE INFOCOM 2004*, Vol. 3. IEEE, 1640–1651.

[70] Anusha Sriraman, Kevin RB Butler, Patrick D McDaniel, and Padma Raghavan. 2007. Analysis of the ipv4 address space delegation structure. In *2007 12th IEEE Symposium on Computers and Communications*. IEEE, 501–508.

[71] Spamhaus Team. 2020. *Suspicious network resurrections*. Retrieved 2021-05-25 from https://www.spamhaus.org/news/article/802/suspicious-network-resurrections

[72] Cecilia Testart, Philipp Richter, Alistair King, Alberto Dainotti, and David Clark. 2019. Profiling BGP Serial Hijackers: Capturing Persistent Misbehavior in the Global Routing Table. In *Proceedings of the Internet Measurement Conference*. 420–434.

[73] Andree Toonk. Dec. 2017. *Popular destinations rerouted to russia*. Retrieved 2020-06-29 from https://bgpmon.net/popular-destinations-rerouted-to-russia/

[74] Ravi Vaidyanathan, Abhrajit Ghosh, Yukiko Sawaya, and Ayumu Kubota. 2012. On the use of enhanced bogon lists (EBLs) to detect malicious traffic. In *2012 International Conference on Computing, Networking and Communications (ICNC)*. IEEE, 1–6.

[75] Q. Vohra and E. Chen. Dec. 2012. *RFC6793: BGP Support for Four-Octet Autonomous System (AS) Number Space*. https://tools.ietf.org/html/rfc6793

[76] Wikipedia. Jun. 2020. *dot-com bubble*. Retrieved 2020-06-29 from https://en.wikipedia.org/wiki/Dot-com_bubble

[77] Rene Wilhelm and Henk Uijterwaal. Oct. 2005. *ASN Missing In Action*. Retrieved 2020-06-29 from https://www.ripe.net/publications/docs/ripe-353

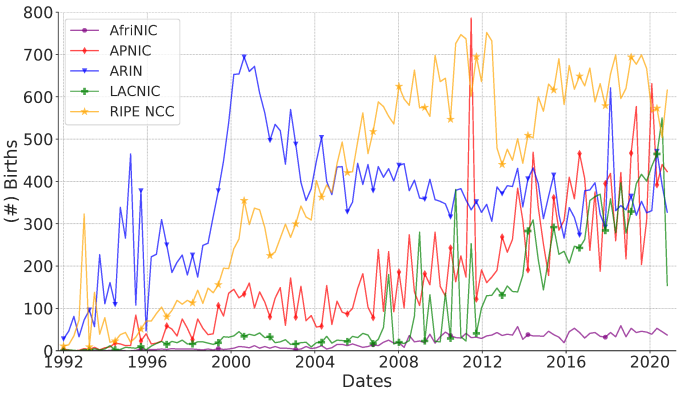

Figure 10: Per-RIR ASN administrative birth rate (3-month bins). It shows the 2000's Internet Bubble and the change in pace of RIPE (2003) and APNIC and LACNIC around 2014

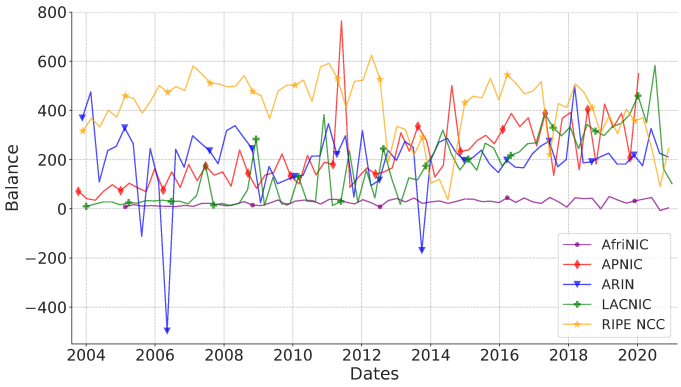

Figure 11: Balance between new ASN allocations and deaths. The volume of RIPE's ASN allocations from 2005 to 2013 is massive. Around 2017, APNIC and LACNIC's ASN allocations exceed ARIN's.

A ADMINISTRATIVE LIFETIME ANALYSIS

In this Appendix, we extend the analysis of §5 based on the administrative and operational ASN lives we build (see §4), providing insights into the expansion of Internet infrastructure in different countries and regions of the world over the years.

Registries growth. When studying ASNs' administrative lives, through the ASN registration date field, we can observe allocations dating back to 1992! In Figure 10, we compare the (quarterly) birth rate of administrative lives across RIRs over time. The graph clearly shows a spike in allocations around year 2000, explainable with the so-called "Internet bubble" [76], and highlights the explosion of LACNIC and APNIC starting from 2014.

Looking at the (quarterly) balance between births and deaths over time (Figure 11) helps us to further capture the infrastructural Internet expansion of these two regions: In the last three years, APNIC and LACNIC have gained more than 1000 ASN net allocations more than ARIN (≈ 4,000 for APNIC and LACNIC and ≈ 3,000 for ARIN). RIPE NCC, still slightly leads, with more than 4,400 ASNs than it had at the beginning of 2018.

Countries infrastructural expansion. The analysis of the ASNs allocations by country, reveals which countries have had faster growth in ASNs allocation in recent years. Brazil is by far the leading country in its region, with an increment in allocations of the total LACNIC ASNs from 64% in 2015 to more than 70% in March 2021 (Argentina is the second country, with only 9.5% of LACNIC ASN allocations). Interestingly, within APNIC, India has climbed to the top (in March 2021, India had more than 15% of all APNIC ASN allocations, while in 2010 it was not even in the top-5!) surpassing Australia, which had been leading in the region since 2006 (Table 4). The third most represented country in the APNIC region is now Indonesia, which recently surpassed China (11.1% and 10.6%, respectively). The ARIN region is dominated by the U.S., with more than 92% of all the allocated resources. In AfriNIC, South Africa is the leading country (with more than 32% of ASN allocations). Finally, in the RIPE region, resources have been distributed more evenly across several countries. Russia largely leads the region with 16.6% of allocated ASNs, more than twice the number of allocated ASNs of the UK, the second largest country.

Table 4: APNIC countries evolution

Pos.	2010	2015	2021
1°	AU: 1038 - 17.6%	AU: 1697 - 16.1%	IN: 2917 - 15.7%
2°	KR: 863 - 14.6%	CN: 1202 - 11.4%	AU: 2681 - 14.5%
3°	JP: 762 - 12.9%	JP: 1103 - 10.4%	ID: 2059 - 11.1%
4°	CN: 449 - 7.6%	IN: 1070 - 10.1%	CN: 1967 - 10.6%
5°	ID: 417 - 7.1%	KR: 1019 - 9.6%	JP: 1127 - 6.1%

16-bit exhaustion. Using our data, we also analyze how close to exhaustion of 16-bit ASNs were the different registries. Looking at the availability of 16-bit numbers, we discover that none of the registries actually used every 16-bit they could allocate. Studying the daily number of 16-bit ASN allocations , the registries reach their maximum in different periods: end of 2013 for AfriNIC, mid-2016 for APNIC, beginning of 2019 for ARIN, mid-2015 for LACNIC, and end of 2018 for RIPE NCC. The global largest number of 16-bit allocations was reached on January 23, 2019, with 60,455 ASNs and globally only 4,039 16-bit available, removing the ones private or reserved by RFC [29, 35, 50].

B RIR POLICIES

The Regional Internet Registries (RIRs) were created in the '90s to manage the delegation of Internet number resources, *i.e.*, Internet Protocol (IP) addresses (IPv4 and IPv6) and AS numbers, at a regional level. Regarding the delegation of Autonomous System Numbers (ASNs), RFC 1930 (also Best Current Practice (BCP) 6) [36] has provided guidelines for the creation and registration of ASNs since it was published in 1996. RFC 1930 has indeed been the baseline of RIR policies for delegating ASNs ever since. The Number Resource Organization (NRO), created in 2003 to coordinate the work of RIRs, has tracked and compared RIR policies—including the ones for allocating AS numbers—since 2004. It publishes the RIR Comparative Policy Overview [55] a few times per year, providing a valuable source about RIR policies and their changes. The next paragraphs describe RIR policies and practices related to the

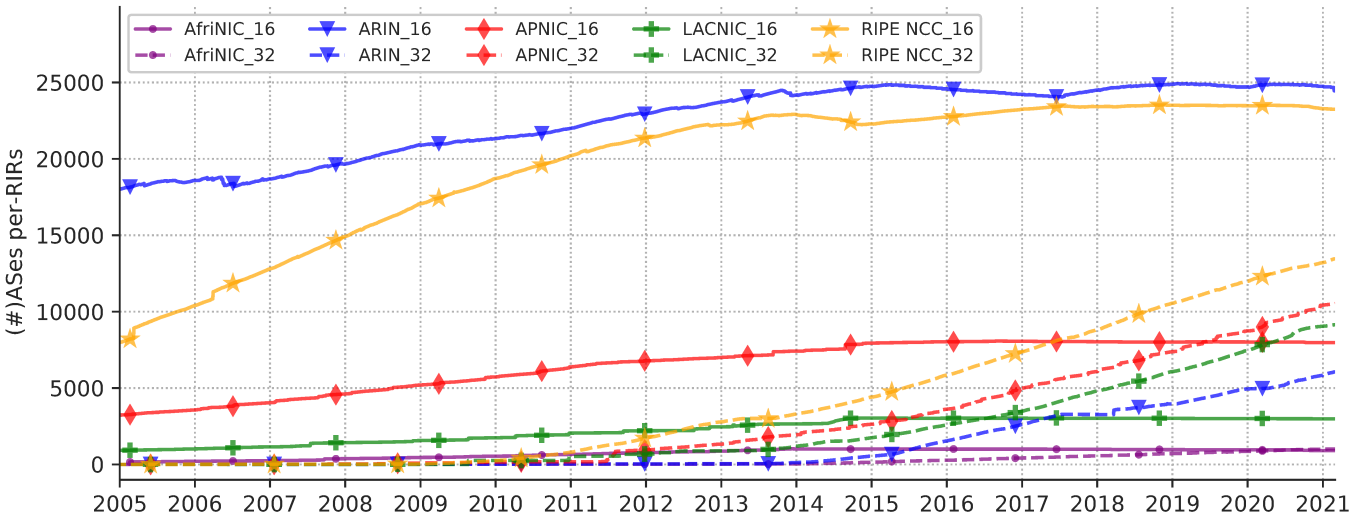

Figure 12: Administrative lives: count of 16-bit (solid lines) and 32-bit (dashed lines) ASNs allocated per day. We can clearly see that the growth of 32-bit allocations is different between the registries. In particular, ARIN 32-bit allocations (dashed blue line) ramp up late (mid-2014) when compared to RIPE NCC, APNIC and LACNIC, despite ARIN being the second-largest registry.

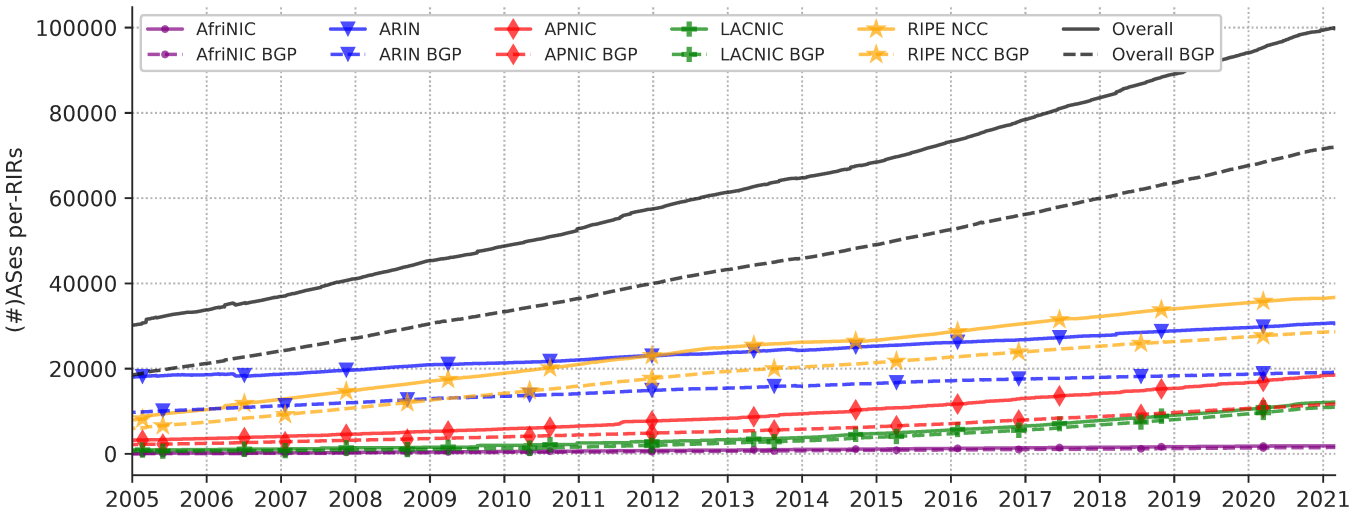

Figure 13: Administrative vs BGP lives: number of ASNs per day that are *administratively* (solid lines) and *operationally* (dashed lines) "alive", per RIR (colored) and overall (black) on a single axis.

allocation of ASNs and how they have changed over time. When possible, we link the allocation process to the delegation files and describe practices related to the tracking of ASN allocations that we infer from our datasets (see §3 for dataset descriptions).

Eligibility Requirements. RIRs have policies that describe which organizations are eligible to be allocated an ASN. In 2004 (the first year with historical policy documents available), ARIN, LACNIC, RIPE NCC and APNIC[14] explicitly cited RFC 1930 in their eligibility criteria, stressing two main conditions:

(1) The organization has a unique routing policy, distinct from its provider (*i.e.,* the provider could not advertise the organization prefixes itself), or
(2) The organization is multihomed.

APNIC used a stricter criteria in 2004, requiring both conditions described above. In addition, APNIC is the only RIR delegating blocks of ASNs to National Internet Registries (NIRs) for further distribution between their members.

Over the years, RIRs have slightly updated the eligibility criteria, mainly to allow organizations to comply with the requirements within 6 months and replacing multihomed with any setting needing to interconnect with an ASN. Starting in 2015, LACNIC also

[14] AfriNIC was being created at the time and defined its policies a few months later.

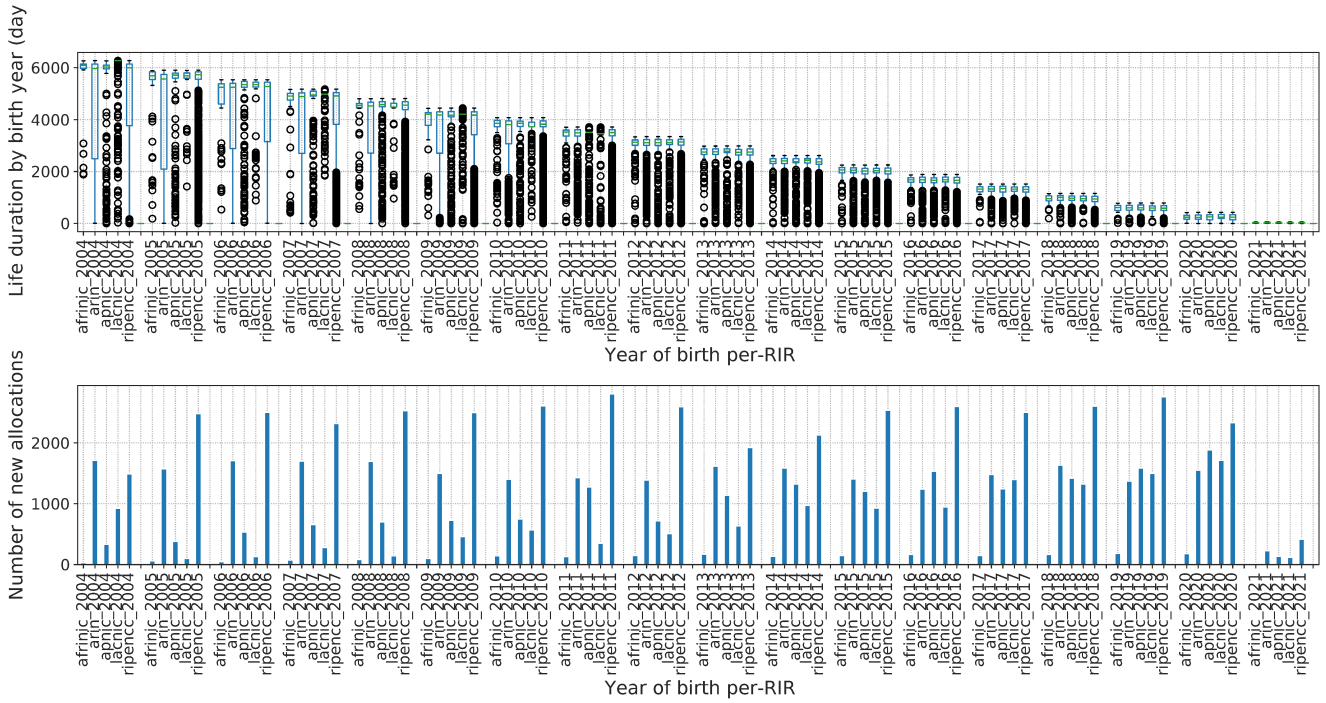

Figure 14: Life expectancy based on year of allocation: The upper sequence of boxplots represents the administrative life duration per registry based on the year of birth (allocation). The bottom image represents the number of new allocations per-RIR for each year. Starting around 2010, the life expectancy becomes similar for all the RIRs, suggesting a kind of life stability.

requests applicants a detailed routing policy, including the list of prefixes they will advertise.

ASN deallocation and reuse. Policies and practices relating to the deallocation of ASNs are only succinctly touched upon in RIRs' policy documents when describing reuse policies. In general, as long as the delegation criteria remains valid, all RIRs will keep a delegation active. However, initially only APNIC had a policy to actively recover unused resources (for the ones it delegated directly, not through a NIR), although all RIRs would put an ASN back in the available pool should the organization the ASN was delegated to cease operations. In 2010, LACNIC and RIPE NCC adopted the policy to actively recover unused resources and ARIN included a policy requesting organizations found to be "materially out of compliance" (*e.g.,* owing the annual fee) to return their resources [11]. Nonetheless, through our exchange with RIRs, we learned that the enforcement of these policies has varied over time. In particular, when 16-bit ASNs became scarce in the mid 2010s, RIPE NCC made the reuse of ASN easier and faster (*e.g.,* not waiting until all dangling announcements of de-allocated ASNs disappear from BGP before putting the ASN back in the available pool). Analyzing the reallocation of AS numbers (reported in Table 2), we identify that indeed ARIN and RIPE have reallocated more resources than the other registries. These practices also impact the deallocation of ASNs and thus the end of the administrative lives we compute in our analysis. We find that it often takes months for AS numbers to be deallocated after their last activity on BGP: the median for APNIC ASNs is more than 6 months, and more than 10 for all the other RIRs.

32-bit ASNs. RIRs started allocating 32-bit ASNs in 2007. At that time, RIRs would delegate 32-bit numbers only if applicants requested 32-bit ASNs. Then, in 2009, RIRs started to delegate 32-bit numbers unless the applicants specifically requested 16-bit ASNs. After this point, RIRs took different paths in the allocation of 16- and 32-bit ASNs. Starting mid 2009, APNIC only allocated 16-bit ASNs if the applicant could "demonstrate that a 32-bit only AS Number is unsuitable". Similarly, in 2010, LACNIC started requesting applicants for 16-bit ASNs to "duly justify the technical reasons" for not using a 32-bit ASN. However, also in 2010, RIPE NCC, ARIN and AfriNIC simply ceased to make any distinction between 16-bit and 32-bit AS Numbers and started assigning them from an undifferentiated 32-bit AS Number allocation pool. In the delegation files dataset, we confirm that RIRs started allocating 32-bit ASN in 2007.[15] Figure 12 shows the number of allocated 16- and 32-bit ASNs per RIR per day. We also notice that the share of allocations 32-bit ASNs represent per RIR evolves differently over time, with the share for APNIC and LACNIC growing much faster than for AfriNIC, RIPE NCC and specially ARIN.

Tracking allocations in delegation files. RIRs use the delegations files to track and make publicly available the allocation records of number resources, including ASNs (for details about the

[15]The one exception is RIPE NCC, which delegated a first 32-bit ASN in December 2006.

Table 5: The table shows how the choice of the inactivity timeout impacts the distribution of cases in our taxonomy.

Inactivity timeout	Complete overlap	Partial overlap	Operational lives outside delegation
15	99,834 (+ 0.04%)	4,390 (- 0.99%)	1,750 (+ 4.9%)
30	99,790	4,434	1,667
50	99,713 (- 0.08%)	4,511 (+ 1.74%)	1,592 (- 4.4%)

content of these files, see §2). Using these files and the methodology described in §4.1, we infer the administrative lifetimes of the ASNs. However, while analyzing the delegation files, we realize that RIRs have different practices when it comes to updating and handling the delegation files. For instance, after allocating an ASN, the precise timing of when the record is added to the file varies. We found that between 90.1% (AfriNIC) and 99.35% (ARIN) of ASN allocations, the ASN appears in the delegation files the same day or the day after its registration. In addition, the outliers that we encountered in our analysis prompted us to exchange emails with RIRs. In §3.1, we describe these phenomena, including the drop of allocated AS numbers, invalid duplicate records, and registration dates that travel back in time. From our exchanges with RIRs, we also learned about challenges they faced in keeping up-to-date the files and dealing with corner cases of resource allocations, both of which sometimes lead to resources disappearing from the files a few days while the issues are being sorted.

C INACTIVITY THRESHOLD

In §4.2 we set a timeout threshold to introduce the concept of operational life of an ASN in BGP. After careful consideration and based on the sensitivity analysis of the distribution of key variables (per-ASN BGP activity gaps and fraction of administrative lives that contain only one or no operational life, shown in Figure 3), we choose a 30 days threshold. Here, to further explore the implications of our choice, we extend our sensitivity analysis to determine how the four categories from our proposed taxonomy (§6) change using either a smaller (15 days) or a larger (50 days) threshold.

In Table 5 we report the impact of 3 different thresholds on the distribution of ASN lives in the 3 categories of our taxonomy that consider operational lives. The highlighted row—the middle row—shows the distribution with a 30-day timeout, the threshold we use in the paper (baseline).The two other rows show numbers for the same categories with the associated threshold (15 and 50), highlighting the delta (in percent) with respect to our 30-day choice. We do not report in the table the never-used category (§6.3) since it is not impacted at all by the choice of the threshold as those ASNs are never seen in BGP.

Table 5 shows that changing the value of the threshold does not have a significant impact on the number of ASNs that completely overlap (§6.1) and partially overlap (§6.2). The most affected category is the "Operational lives without allocation" (§6.4). However, it is a small fluctuation of less than 5%, that is almost symmetric around the threshold we picked. These changes are not significant and do not alter the substance of our findings.