BGP Zombies: an Analysis of Beacons Stuck Routes

Romain Fontugne, Esteban Bautista, Colin Petrie, Yutaro Nomura, Patrice Abry, Paulo Gonçalves, Kensuke Fukuda, Emile Aben

BGP Zombies: an Analysis of Beacons Stuck Routes

Romain Fontugne[1], Esteban Bautista[2], Colin Petrie[3], Yutaro Nomura[4], Patrice Abry[5,6], Paulo Goncalves[2], Kensuke Fukuda[6], and Emile Aben[3]

[1] IIJ Research Lab, Tokyo, Japan romain@iij.ad.jp
[2] Univ Lyon, Ens de Lyon, Inria, CNRS, UCB Lyon 1, F-69342, Lyon, France
[3] RIPE NCC, Amsterdam, Netherlands
[4] The University of Tokyo, Tokyo, Japan
[5] Univ Lyon, Ens de Lyon, Univ Claude Bernard, CNRS, Laboratoire de Physique, Lyon, France
[6] NII / Sokendai, Tokyo, Japan

Abstract. Network operators use the Border Gateway Protocol (BGP) to control the global visibility of their networks. When withdrawing an IP prefix from the Internet, an origin network sends BGP withdraw messages, which are expected to propagate to all BGP routers that hold an entry for that IP prefix in their routing table. Yet network operators occasionally report issues where routers maintain routes to IP prefixes withdrawn by their origin network. We refer to this problem as BGP zombies and characterize their appearance using RIS BGP beacons, a set of prefixes withdrawn every four hours.Across the 27 monitored beacon prefixes, we observe usually more than one zombie outbreak per day. But their presence is highly volatile, on average a monitored peer misses 1.8% withdraws for an IPv4 beacon (2.7% for IPv6). We also discovered that BGP zombies can propagate to other ASes, for example, zombies in a transit network are inevitably affecting its customer networks. We employ a graph-based semi-supervised machine learning technique to estimate the scope of zombies propagation, and found that most of the observed zombie outbreaks are small (i.e. on average 10% of monitored ASes for IPv4 and 17% for IPv6). We also report some large zombie outbreaks with almost all monitored ASes affected.

1 Introduction

BGP is the protocol that governs inter-domain routing on the Internet. As such understanding the boundaries of its behaviour is of prime importance. The tens of thousands of Autonomous Systems (ASes) that constitute the Internet expect to rapidly be able to change the routing and reachability of the address space they are originating towards all other ASes. The process of announcing and withdrawing address space is of utmost importance.

When an origin AS withdraws a prefix, it sends a withdrawal message to its BGP neighbours, who will in turn propagate it to their neighbours. Sometimes a network sees the best path that it propagated to neighbours disappears, but in a

rich topology the network still has alternative paths yet to be withdrawn. In that case the neighbours will not receive a withdrawal, but the best alternative path. This process, called *path hunting*, typically causes several BGP path changes in the matter of minutes, before a BGP prefix is fully withdrawn [9]. The richer the topology between the origin AS and a BGP speaker, the larger the number of path changes.

Theoretically this withdrawal process ends with the prefix completely withdrawn from all BGP speakers, as announcements and withdrawals propagate through the entire Internet similarly. In practice, this sometimes fails, a phenomenon known by network operators as *stuck routes* or *zombie routes*. In this case, path hunting gets stuck in a state where BGP routes are still visible at some BGP routers, something we can easily observe with route collector systems like RIS, Routeviews, and Isolario [6,7,2].

This work is motivated by the operational confusion that missing withdrawal causes. We have witnessed several cases where zombie routes caused confusion about the state of the withdrawn address space. In addition, troubleshooting and cleaning zombie routes is a burden for network operators. This phenomenon is relatively unknown outside network operator circles, and generally not well understood. We intend to shed light on BGP zombies in order to make the research community aware of this problem and to assist operators.

In this study we characterize zombie routes in a controlled setting using the RIS routing beacons. In this controlled environment, we can measure the frequency of failed withdrawals, and alternative paths that are seen in the withdrawal phase. The key contributions of this paper are to provide the first characterization of BGP zombies and a method to infer the scope of zombie outbreaks with the help of a graph-based semi-supervised machine learning algorithm. Our experiments reveal a surprisingly high number of zombies. Zombies are seen daily in our dataset, but we found that the number of affected ASes is usually limited (on average 10% of monitored ASes in IPv4 and 17% for IPv6). The appearance of zombie routes is very erratic. Zombie routes rarely emerge for numerous prefixes at the same time and for the same RIS peers. The average likelihood of observing a zombie for a given RIS peer and beacon prefix is 1.8% for IPv4 and 2.7% for IPv6. Finally, we show that numerous zombie paths are revealed during path hunting and the scope of an outbreak is usually related to the affected transit networks.

2 BGP zombies

Before diving into the detailed analysis of BGP zombies, we define all the related terminology and explain our experimental setup. A **BGP zombie** refers to an active Routing Information Base (RIB) entry for a prefix that has been withdrawn by its origin network, and is hence not reachable anymore. In this paper we also refer to **zombie ASes** and **zombie peers** for ASes and BGP peers whose routers have BGP zombies. We refer to all zombies that correspond

to the same prefix and appear during the same two-hour time slot as a **zombie outbreak**, the outbreak size is the number of zombie ASes.

2.1 Experimental setup

In order to observe BGP zombies one needs to withdraw an IP prefix from its origin AS and inspect RIB changes, or lack thereof, in other ASes. We conduct such controlled experiments with the help of RIPE's Routing Information Service (RIS) BGP beacons [14,4] and RIS BGP data repository [6].

The RIS BGP beacons are a set of IPv4 and IPv6 prefixes that are used solely for studying Internet inter-domain routing. These IP prefixes are announced and withdrawn at predetermined time intervals. Namely, RIS BGP beacons are announced every day at 00:00, 04:00, 08:00, 12:00, 16:00, and 20:00 UTC, and they are withdrawn two hours after the announcements (i.e. at 02:00, 06:00, 10:00, 14:00, 18:00, and 22:00 UTC). We are monitoring 27 beacon prefixes (13 IPv4 and 14 IPv6) announced from Europe, U.S.A., Russia, Japan, and Brazil.

RIS also archives RIB and BGP update messages collected at diverse places on the Internet. RIS collectors (named rrc00, rrc01, etc...) are mainly located at Internet eXchange Points (IXP) and peer with hundreds of different ASes. Using this archive we can monitor how these ASes respond to the BGP beacons stimuli and characterize the emergence of BGP zombies.

For beacon prefixes, the detection of zombies in RIS peers is straightforward. We keep track of the visibility of beacons for all RIS peers and report a zombie for each RIB entry that is still active 1.5 hour after the prefix was withdrawn. The 1.5 hour delay is set empirically to avoid late withdrawals due to BGP convergence [14], route flap damping [20], or stale routes [17]. Each beacon's visibility is monitored in near-real time using the RIPEstat looking glass [5] so we can trigger active measurements (e.g. traceroutes) during detected zombie outbreaks.

We conducted experiments during the three periods of time listed in Table 1 and detected for the 27 monitored prefixes a total of 5115 zombie outbreaks, each composed of one or more zombie routes for the same prefix.

Table 1. Measurement periods and number of detected zombie outbreaks for the 27 monitored beacons.

Start	End	#IPv4 outbreaks	#IPv6 outbreaks
2017-03-01	2017-04-28	1732	591
2017-10-01	2018-12-28	384	1202
2018-07-19	2018-08-31	520	686

2.2 Example

Figure 1 illustrates the visibility for beacon 84.205.71.0/24 from all RIS peers on September 9th and 10th, 2017. Peers are sorted on the y axis and time is represented by the x axis. From 12:00 to 18:00 UTC, all peers behave as expected. At 12:00, RIS peers announce the availability of the beacon prefix and maintain an

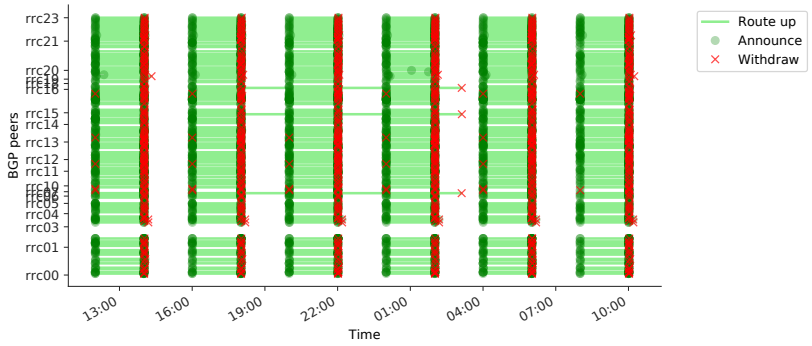


Fig. 1. Visibility for 84.205.71.0/24 from all RIS collectors on September 9th and 10th, 2017. A zombie outbreak happened from 18:00 to 20:00 UTC and another one from 22:00 to 00:00 UTC. Both outbreaks are visible from three RIS peers.

active route to the prefix until 14:00. One peer from rrc19 withdraws the prefix a bit late (14:19), but this is not considered as a zombie because the prefix is withdrawn reasonably quickly. However, at 18:00 three peers do not withdraw the beacon although this prefix is not reachable at that time. This zombie outbreak ends at 20:00 when the beacon is re-announced. A similar zombie outbreak appears at 22:00 for the same three peers.

During the first zombie outbreak (18:00-20:00), we found other zombies for the same three peers but another beacon (84.205.67.0/24). The 25 other beacons are withdrawn as expected at that time. For the second outbreak (22:00-00:00), we found no other zombie. These observations give an early glimpse of the relationship between outbreaks for different prefixes. Zombie outbreaks for different beacons can be related but are usually independent. We formally investigate the co-occurrence of outbreaks from different beacons in Section 4.1.

2.3 Are zombies real?

To ensure that no artificial zombies are caused by measurement artifacts, we also looked for zombie evidences in other datasets.

First, for each zombie detected with the RIPEstat looking glass, we also accessed the raw data from the RIS archive using BGPstream [16] and checked that the withdraw messages are indeed missing in the raw traces. We found 794 outbreaks that are reported by the looking glass but not present in the raw data. We ignored these events in our analysis; these are not listed in Table 1.

Then, we also looked at the presence of zombies in Routeviews data and NLNOG looking glass during large zombie outbreak and confirmed that zombies are also present there. As Routeviews and RIS are now using completely different software for data collection (ExaBGP vs. Quagga/Zebra) we assume that observed zombies are not caused by malfunctioning collectors.

Finally, during zombie outbreaks we performed traceroute measurements towards beacon prefixes from Atlas probes located in zombie ASes. The traceroutes

reveal that border routers in zombie ASes are indeed forwarding packets whereas other routers usually drop these packets. We also use these traceroute results to evaluate our method to infer zombie ASes on AS paths (Section 3.2).

3 Hunting zombies

With the simple zombie detection technique described above, we observe zombies only in ASes that are peering with RIS collectors. In this section, we show that the withdrawn and zombie AS paths collected by RIS also enable us to infer zombie ASes beyond RIS peers and estimate the scope of outbreaks.

For each outbreak we retrieve the AS path of zombie entries and the last valid path for peers that have correctly withdrawn the beacon. A path alone provides little information, but put together they reveal topological similarities that we consider evidence for the locations of zombies.

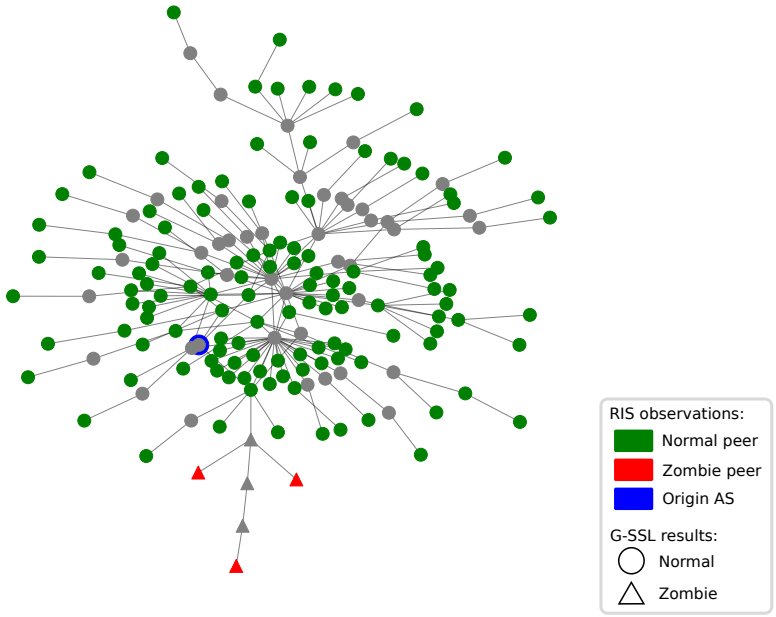


Fig. 2. AS paths for the second outbreak in Fig.1. Each node is an AS, red and green nodes are RIS peers. Gray nodes are ASes seen on the paths but not peering with RIS.

Figure 2 depicts AS paths for the second outbreak in Figure 1. Each node represents an AS and consecutive ASes in the AS paths are connected by an edge. The green nodes represent RIS peers that have correctly withdrawn the prefix at 22:00. The red nodes represent zombie peers observed from 22:00 to 00:00. The gray nodes represent ASes that are not peering with RIS collectors, hence we have no direct observations for these ASes though they appear in collected AS paths. Here, the three observed zombies share the same upstream provider which

is strong evidence that this provider and all its downstream ASes (depicted by triangles in Figure 2) are also zombies.

To systematically identify these clusters of zombies, we build such graphs for each outbreak then we classify unknown ASes using the graph-based machine learning technique described in the next section. The results of the classification are illustrated in Figure 2 with the shape of the nodes: triangles represent detected zombies; circles represent other ASes.

3.1 Graph-based Semi-Supervised Learning

Graph-based Semi-Supervised Learning (G-SSL) is a generic framework permitting efficient classification of graph nodes by jointly exploiting the graph topology and prior information consisting of a small fraction of nodes being a priori classified by *experts* [19] (i.e. RIS peers). There already exist several documented examples where G-SSL has outperformed other state-of-the-art classification strategies (e.g., BitTorrent content and user classification [10], text recognition [18], bio-medical diagnoses [21]).

Amongst the several versions of G-SSL, the PageRank-based G-SSL is a popular and commonly used one [11]. It relies on a coding of the graph topology via a specific operator, the (combinatorial) Laplacian L. Namely, let us consider an N node undirected graph encoded by the adjacency matrix W, with $W_{i,j} = 1$ when nodes i and j are connected and 0 otherwise. Further, let $d_i = \sum_j W_{ij}$ denote the degree of node i, $D = diag(d_1, \ldots, d_N)$ the diagonal matrix of vertex degrees, and form $L = D - W$. The PageRank K-class classification procedure can be sketched as follows. The labeled information is encoded in a matrix $Y \in \mathbb{R}^{N \times K}$, where $Y_{ik} = 1$ if node i is declared by expert to belong to class k and 0 elsewhere. In the present work, Y conveys the information provided by RIS; normal and zombie peers are respectively coded as $Y_{i1} = 1$ and $Y_{i2} = 1$. The classification of the unlabeled nodes amounts to estimate a vectorial signal $X \in \mathbb{R}^{N \times K}$ on the graph as:

$$\min_x \left\{ x^T D^{-1} L D^{-1} x + \mu \left(y - x\right)^T D^{-1} \left(y - x\right) \right\}. \tag{1}$$

This functional minimization is known to have an analytical closed-form solution, providing access to X, without recourse to a time/memory consuming iterative minimization procedure:

$$X^T = \frac{\mu}{\mu + 2} \mathbf{y}^T \left(\mathbb{I} - \alpha \mathbf{D^{-1}W} \right)^{-1}. \tag{2}$$

Once X is computed, node i is assigned to the class k selected by $\text{argmax}_k X_{ik}$.

The hyper-parameter μ balances the confidence granted to the *expert knowledge* versus the information conveyed by the graph (and the graph Laplacian L). It is tuned by means of a standard leave-one-out cross validation procedure, tailored to the context of semi-supervised learning: From the set of documented vertices, one element, per class, is selected as a labeled example, while the rest is added to the group of not documented and used for validation. The procedure is repeated and μ is selected as maximizing average detection performance.

3.2 Validation

G-SSL produces a list of zombie ASes that are not necessarily peering with RIS collectors. To evaluate the classification accuracy of G-SSL we performed timely traceroute measurements from ASes found on the zombie paths and compared the traceroute results with G-SSL results.

Our traceroute measurements are done with the RIPE Atlas measurement platform [3]. We select five Atlas probes for each AS found in zombie paths, and perform traceroutes towards the corresponding beacon prefix every 5 minutes until the prefix is announced again.

Comparing traceroute results to G-SSL results requires certain precautions. We intuitively expect routers from zombie ASes to forward traceroute packets and other routers to either drop these packets or return an ICMP network unreachable error. However, the presence of default routes in intra-AS routing is inevitably exhibiting router IP addresses although the AS border routers have withdrawn the prefix. Another difficulty is to identify borders between two ASes and avoid making wrong inferences when mapping IP addresses to AS numbers [15,13].

To address both issues we employ the following heuristics. First, we discard the first public IP found in traceroutes as it usually stands for a gateway with a default route. We group all traceroutes initiated from the same AS, if these traceroutes consist only of ICMP network unreachable errors and unresponsive routers then we consider that AS as normal, that is the AS has correctly withdrawn the route and is not forwarding packets. For traceroutes with responsive routers we retrieve the routers' ASN using longest prefix match and compute F_A, the number of IP addresses from ASN A that forwarded packets, and, E_A the number of IP addresses from ASN A that sent an ICMP error. We consider an AS A as zombie if the majority of its routers are forwarding packets, $F_A > E_A$.

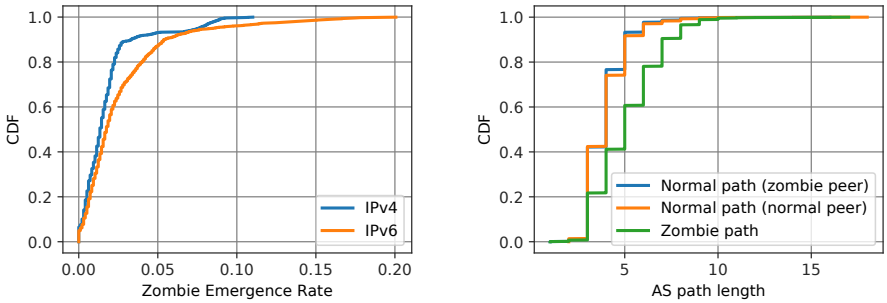
The AS classification using traceroutes and the observations from RIS peers constitute the ground truth data we use to evaluate G-SSL results. For the three measurement periods G-SSL retrieved 97% of the zombies identified in the ground truth and 99% of the normal AS, which is more than acceptable for the following characterization of zombies. Since G-SSL classifies all nodes in the graph, we also obtain 35% more classified ASes than using traceroutes and RIS peers.

4 Zombie characteristics

We now investigate temporal and topological characteristics of zombies directly observed at RIS peers and those inferred using the G-SSL method. Our aim here is to quantify the frequency of zombies, uncover their locality, and estimate the scale of zombie outbreaks.

4.1 Zombies observed at RIS peers

Starting with zombies observed at RIS peers, we compute the zombie emergence rate, that is the number of times zombies are reported for each peer and each



(a) Frequency of zombie appearance for each RIS peer and beacon prefix.

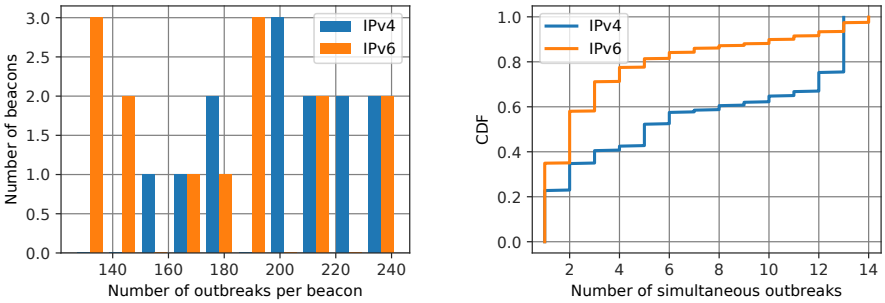
(b) AS path length for IPv4 beacons.

Fig. 3. Zombies observed by RIS peers.

beacon normalized by the number of times beacons have been withdrawn during our measurement study. This metric corresponds to the likelihood of pair $\langle peer, beacon \rangle$ to cause a zombie. Figure 3a depicts the distribution of the values obtained with our dataset. We observe only 6.5% $\langle peer, beacon \rangle$ pairs with no zombie during our entire measurement periods. However, zombies are uncommon for RIS peers, 50% of the $\langle peer, beacon \rangle$ pairs have zombie entries for less than 1.3% of the beacon withdraws (average value is 1.8% for IPv4 and 2.7% for IPv6). We found some outlier values, meaning that a few RIS peers are more prone to zombies, which is better understood with G-SSL results (Section 4.2).

We also compared the zombie AS paths to the paths that are advertised before the beacon withdraw. For IPv4, 50% of the zombie paths are different than the paths that are used before the withdraw (69% for IPv6). Figure 3b illustrates the distribution of path length for zombie paths, paths that were previously advertised by zombie ASes (Normal path (zombie peer)), and paths that were advertised by peers that correctly withdrawn the beacon (Normal path (normal peer)). The distribution of zombie paths is clearly shifted to the right hence zombie paths are usually longer. These observations imply that zombie paths are mostly different from the paths that are selected during BGP path convergence, and numerous zombies appear during path hunting.

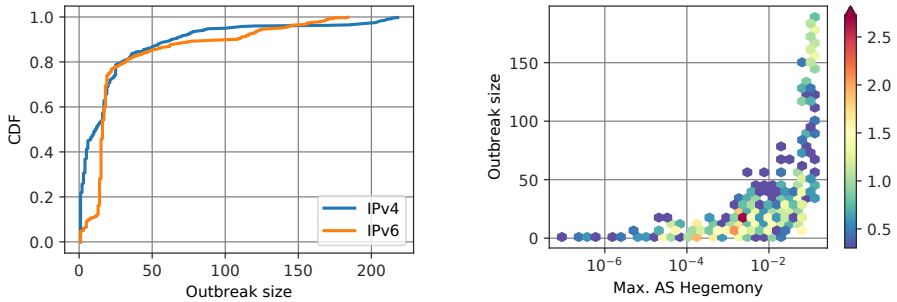
Then we examine if certain beacons are more prone to zombies. Figure 4a shows the number of zombie outbreaks observed per beacon. On average we detect about 200 outbreaks per beacon in our dataset. For IPv6 beacons announced from DE-CIX in Frankfurt and VIX in Vienna are responsible for the largest number of outbreaks. For IPv4 the beacon with the most outbreaks is the one announced from both AMS-IX and NL-IX in Amsterdam. To understand the relationship between zombies detected across the various beacons, we compute the number of outbreaks that happened simultaneously but for different beacons. For 23% of instances where we detect IPv4 zombies (35% for IPv6) we found zombies only for a single beacon. For IPv4 we also found multiple



(a) Total number of zombie outbreaks per beacon.

(b) Number of simultaneous zombie outbreaks.

Fig. 4. Dependency of outbreaks across BGP beacons.



(a) Distribution of the number of zombie ASes per outbreak.

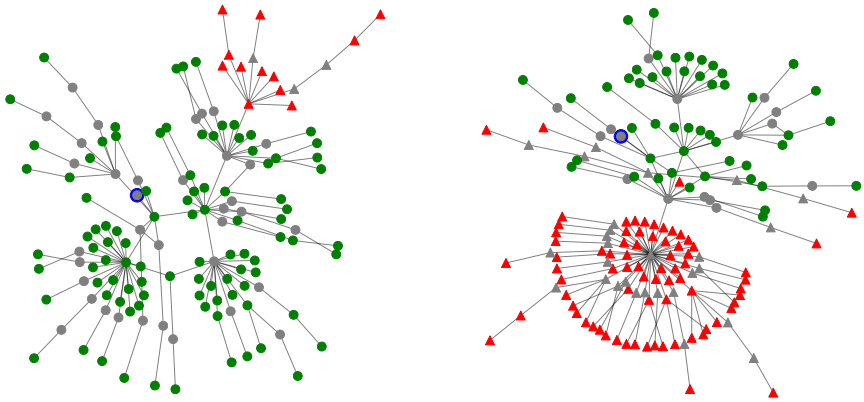
(b) Relation between main zombie transit and outbreak size.

Fig. 5. All detected Zombies (i.e. observed by RIS and inferred by G-SSL).

instances (25%) where we detect simultaneous zombies outbreaks for all monitored beacons. The rest of the distribution is uniform, meaning that we observe little correlation between outbreaks on different beacons. These observations reveal that usually outbreaks emerge independently across different prefixes, yet in certain cases some peers altogether miss withdraws for all monitored beacons.

4.2 Zombies beyond RIS peers

Using G-SSL results we can explore the scale of zombie outbreaks beyond the monitored RIS peers. For each zombie outbreak we count the total number of ASes with detected zombies (i.e. zombies observed at RIS peers and zombies inferred by G-SSL). On average, a zombie outbreak affects 24 ASes for IPv4 and 30 ASes for IPv6, that is 10% of the IPv4 monitored ASes and 17% for IPv6.



(a) Zombie detected in Init7 for beacon 2001:7fb:fe06::/48 on March 1st, 2017.

(b) Zombie detected in Level(3) for beacon 84.205.70.0/24 on December 6th, 2017.

Fig. 6. Examples of zombie outbreak affecting significant transit networks. See Figure 2 for the legend.

However, the distribution of outbreak size is significantly skewed (Fig. 5a). The median outbreak size is 11 ASes for IPv4 and 16 ASes for IPv6. We also observe a few instances where most of the monitored ASes are zombies due to zombies that appeared close to the beacons' origin AS or in large ISPs.

For IPv6 we found that a remarkably high number of outbreaks (63%) contain between 12 and 19 ASes. For IPv4, the number of outbreaks with that particular size is also significant (18%), but we also observe a large proportion of smaller outbreaks, 45% of the IPv4 outbreaks have between 1 and 6 ASes.

By manually looking at the results we noticed certain patterns among outbreaks. We hypothesize that the number of zombie ASes is usually related to the importance of the transit networks affected by zombies. To illustrate this we select for each outbreak the most prominent transit network affected by the outbreak using global AS hegemony [8,12]. AS hegemony measures the centrality of an AS in the Internet, higher values standing for Tier-1 ISPs. Comparing the size of outbreaks to the largest AS hegemony score of affected ASes (Fig. 5b) shows that small outbreaks consist only of edge networks (i.e. low AS hegemony) and large transit networks belong only to the largest outbreaks.

Figure 6 illustrates two outbreaks where we detected zombies in large transit networks. The left hand side graph (Fig. 6a) represents an outbreak where the zombie AS with the highest hegemony score is Init7 and all ASes downstream are also affected by the outbreak. The right hand side graph (Fig. 6b) depicts another outbreak where we inferred a zombie in a Tier-1 network, Level(3). As Level(3)'s customer cone is larger the scope of the outbreak is also more important. This results in about half of the RIS peers having zombie routes through Level(3).

Table 2. Top 5 affected transit ASes for IPv4, IPv6, and each measurement period. Each percentage is the number of outbreaks that include the AS divided by the total number of outbreaks for the corresponding measurement period.

(a) IPv4

Mar./Apr. 2017		Oct./Dec. 2017		Jul./Aug. 2018	
AS3303 Swisscom	46.13%	AS6939 HE	14.84%	AS6667 Elisa	19.81%
AS12874 Fastweb	46.07%	AS1103 SURFnet	9.90%	AS680 DFN	17.69%
AS8359 MTS	9.93%	AS7575 AARNet	9.38%	AS7018 AT&T	16.73%
AS680 DFN	9.18%	AS286 KPN	9.38%	AS3549 Level3 GBLX	15.96%
AS7018 AT&T	8.60%	AS6453 TATA	9.11%	AS7575 AARNet	15.19%

(b) IPv6

Mar./Apr. 2017		Oct./Dec. 2017		Jul./Aug. 2018	
AS8455 Atom86	39%	AS13030 Init7	57%	AS13030 Init7	74%
AS13030 Init7	39%	AS8455 Atom86	55%	AS8455 Atom86	73%
AS5580 Hibernia	36%	AS8928 Interoute	36%	AS7018 AT&T	15%
AS7018 AT&T	8%	AS9002 RETN	35%	AS23106 CEMIG	13%
AS28917 Fiord	6%	AS33891 Core-Backbone	22%	AS1916 RNP	13%

In the absence of zombies we observe much less AS paths that contain Init7 or Level(3). This demonstrates again the role of path hunting in zombie propagation. When a beacon is withdrawn and a zombie appears on a transit network, downstream ASes are selecting that zombie path as other paths get discarded.

The frequency of zombies at transit networks is hence directly related to the topological spread of zombie outbreaks reported earlier (Fig. 5a). In Table 2 we list transit networks that appeared the most in zombie outbreaks. We again employ AS hegemony to focus only on large transit ASes, we arbitrarily picked ASes with an hegemony higher than 0.001. For IPv4 the top-5 ASes vary significantly across the three measurement periods. For IPv6 we found that Init7 and Atom86 are always the top two affected networks. Our manual inspection of the data reveals that Atom86 is downstream of Init7, so is affected every time Init7 has zombies. Init7's zombies usually propagate to 14 downstream ASes (example shown in Figure 6a), which explains the large number of outbreaks composed of about 15 ASes in IPv6 in our results (Fig. 5a).

Network operators at Init7 acknowledged these issues with IPv6 routes, likely due to misbehaving vendor software, and expressed the need for zombie reporting systems, as it creates customer complaints every few months. Mitigation of the BGP zombies usually required the clearing of some Route Reflector iBGP sessions within Init7's network. Init7 operates its backbone using Extreme Networks MLXe (formerly known as Brocade MLXe) platform, which seems to be uncommon. Upgrading to later firmware version did not resolve the problem. Notice that we do not imply that detected outbreaks are caused by the transit networks listed in Table 2. Finding the root cause of zombie outbreaks requires additional measurements within these networks and their peers.

5 Discussion

While detecting BGP zombies with RIS beacons is straightforward, we faced significant challenges in pinpointing the root cause of observed zombies. Given the erratic patterns observed in our study and the investigations conducted with network operators, we believe zombies are mainly the results of software bugs in routers, BGP optimizers, and route reflectors. The systematic identification of zombie root causes on the Internet has proven to be very challenging, even for operators, as it requires timely and detailed information from a complex and occasionally misbehaving infrastructure. It is however a crucial challenge to ensure that this issue will not cause an increasing amount of difficult to debug issues for network operators.

If the fraction of zombie routes in the wild is in the same order of magnitude as what we see for RIS beacons, this can have interesting consequences that would merit further research. For instance, in the case of large route leaks, zombie routes could add significantly to the complexity of mitigating these incidents.

Our study focuses only on RIS beacons as we know their withdraw times a priori. However, these results cannot be easily extrapolated for any routed prefix. We could infer zombies for cases where a prefix is withdrawn in a short period of time for most, but not all route collector peers, and it remains difficult to distinguish this from a routing configuration change intended to limit the visibility of a prefix. Furthermore, in the case of large zombie outbreaks, which are of prime interest, one may confuse the few observed withdraws with a local routing issue. We plan to address these challenges in future works. A rigorous method for detecting zombies in the wild would allow us to estimate the overall impact of zombies on routing tables and to provide network operators with tools to effectively identify zombies.

6 Conclusions

In this paper, we investigated the emergence of BGP zombies with the help of RIS beacons. Our study spans across a year and half of data and revealed that BGP zombies are seen daily, although the scope of outbreaks is usually limited to a small fraction of monitored ASes (on average 10% for IPv4 and 17% for IPv6). We found almost no regularity in the appearance of zombies. They rarely emerge synchronously on all monitored prefixes. Numerous zombie paths are revealed during path hunting and the scope of an outbreak is usually related to the affected transit networks. Our future plans are to identify zombies for any prefix announced on the Internet (i.e. not only beacon prefixes) and quantify the impact of zombies in the wild. Finally, we make our tools and traceroute results publicly available [1] in order to share our findings and assist researchers in their zombie hunt.

References

1. BGP Zombie: tools and data. https://github.com/romain-fontugne/BGPzombiesSSL.
2. Isolario project. https://www.isolario.it/.
3. RIPE NCC, Atlas. https://atlas.ripe.net.
4. RIPE NCC, Current RIS Routing Beacons. https://www.ripe.net/analyse/internet-measurements/routing-information-service-ris/current-ris-routing-beacons.
5. RIPE NCC, RIPEstat: BGP Looking Glass. https://stat.ripe.net/widget/looking-glass.
6. RIPE NCC, RIS Raw Data. https://www.ripe.net/analyse/internet-measurements/routing-information-service-ris/ris-raw-data.
7. The RouteViews project. http://www.routeviews.org/.
8. AS Hegemony Results. http://ihr.iijlab.net/ihr/hegemony/, 2017.
9. V. Asturiano. The Shape of a BGP Update. https://labs.ripe.net/Members/vastur/the-shape-of-a-bgp-update.
10. K. Avrachenkov, P. Gonçalves, A. Legout, and M. Sokol. Classification of content and users in BitTorrent by semi-supervised learning methods. *8th International Wireless Communications and Mobile Computing Conference (IWCMC)*, pages 625–630, Aug 2012.
11. K. Avrachenkov, A. Mishenin, P. Gonçalves, and M. Sokol. Generalized Optimization Framework for Graph-based Semi-supervised Learning. *Proceedings of the 2012 SIAM International Conference on Data Mining*, pages 966–974, 2012.
12. R. Fontugne, A. Shah, and E. Aben. The (thin) Bridges of AS Connectivity: Measuring Dependency using AS Hegemony. In *Proceedings of PAM'18*. LNCS, 2018.
13. M. Luckie, A. Dhamdhere, B. Huffaker, D. Clark, and k. claffy. Bdrmap: Inference of borders between ip networks. In *Proceedings of the 2016 Internet Measurement Conference*, IMC '16, pages 381–396, New York, NY, USA, 2016. ACM.
14. Z. M. Mao, R. Bush, T. G. Griffin, and M. Roughan. Bgp beacons. In *Proceedings of the 3rd ACM SIGCOMM conference on Internet measurement*, pages 1–14. ACM, 2003.
15. A. Marder and J. M. Smith. Map-it: Multipass accurate passive inferences from traceroute. In *Proceedings of the 2016 Internet Measurement Conference*, IMC '16, pages 397–411, New York, NY, USA, 2016. ACM.
16. C. Orsini, A. King, D. Giordano, V. Giotsas, and A. Dainotti. Bgpstream: a software framework for live and historical bgp data analysis. In *IMC*, pages 429–444. ACM, 2016.
17. S. Sangli, E. Chen, C. Systems, R. Fernando, J. Scudder, and Y. Rekhter. Graceful Restart Mechanism for BGP (No. RFC 4724). Technical report, 2007.
18. A. Subramanya and J. Bilmes. Soft-supervised learning for text classification. *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 1090–1099, 2008.
19. A. Subramanya and P. P. Talukdar. *Graph-Based Semi-Supervised Learning*. Morgan & Claypool Publishers, 2014.
20. C. Villamizar, R. Chandra, and R. Govindan. BGP route flap damping (No. RFC 2439). Technical report, 1998.
21. M. Zhao, R. H. M. Chan, T. W. S. Chow, and P. Tang. Compact graph based semi-supervised learning for medical diagnosis in alzheimer?s disease. *IEEE Signal Processing Letters*, 21(10):1192–1196, Oct 2014.