This is the author's version of an article that has been published in IEEE. Changes were made to this version by the publisher prior to publication:

Garcia-Martinez, A., Bagnulo, M. (2019). Measuring BGP Route Propagation Times. *IEEE Communications Letters*, 23(12), 2432–2436.

DOI: <http://dx.doi.org/10.1109/LCOMM.2019.2945964>

Copyright © 2019 IEEE. Personal use is permitted. For any other purposes, permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.

Measuring BGP route propagation times

Alberto García-Martínez and Marcelo Bagnulo

Abstract—We present a methodology to measure the propagation time of routes with BGP route collectors and beacons which accounts for the clock offset error between route source and destinations. We apply it to the RIPE RIS data set comparing 2012 and 2018, showing that the IPv4 route prefix reachability interval and the time to receive the preferred route have not changed much, but the time to remove an IPv4 route has increased and it is high (a mean of 121s to remove a route). In 2018, IPv6 route propagation times are shorter than IPv4 ones.

Index Terms—BGP, Route propagation, MRAI, BGP beacons, Route collectors

I. INTRODUCTION

BGP is the routing protocol that glues together the networks forming the Internet. Its ability to timely react to connectivity or policy changes is critical for the quality of the communications. The reaction time of BGP is influenced, among other elements, by the BGP route propagation delay, i.e., the time between a route announced in the origin router and it is received by the other routers. The BGP route propagation delay is determined by many factors, including:

- **Number of routers traversed.**
- **Time to process the route** in each of the traversed routers, depending on the router processing capacity and on the number of pending routes to process.
- **Minimum Route Advertisement Interval (MRAI).** The MRAI timer specifies a delay of the propagation of BGP updates from router to router in order to reduce the number of BGP messages exchanged, and thus the processing requirements. The BGP specification [1] defines a default MRAI value of 30s for routes sent to a different Autonomous System (AS), and 5s for routes sent to routers of the same AS.
- **Route Flap Damping (RFD).** RFD [2] is a mechanism by which route propagation is delayed if the router detects a high number of advertisements related with the same prefix in a period of time, as the result of a penalty that increases every time a route for a prefix is received. It aims to reduce route instability and persistent oscillations.

Manuscript received July XX, 2019; accepted September XX, 2019. Date of publication XXXXXXXX XX, 2019; date of current version October X, 2019. The work of Alberto García-Martínez was supported by 5G-Coral project, H2020-ICT-2016-2. The work of Marcelo Bagnulo was supported in part by the Spanish Ministry of Economy and Competitiveness under 5GCity project, TEC2016-76795-C6-3-R, and in part by the European Commission, Marie Curie ITN, METRICS project. The associate editor coordinating the review of this letter and approving it for publication was XXXXXXXXXX (Corresponding author: Alberto García-Martínez.)

The authors are with the Department of Telematics Engineering, Universidad Carlos III de Madrid, 28911 Leganés, Madrid, Spain (e-mail: alberto@it.uc3m.es)

Digital Object Identifier 10.1109/LCOMM.2019.XXXXXXX

Previous efforts to measure the route propagation delay in BGP were performed in 2001 by Labovitz et al. [3]. They induced route updates and withdraws at random instants and analyzed the traces observed in ROUTEVIEW collectors. NTP was used to synchronize the clocks between the systems generating and collecting the routes. They indicated that the route propagation time for route withdrawn could be approximated as the number of ASes in the longest propagation path by 15s (half the MRAI default value).

Since 2001, many things have changed in the Internet ecosystem. Besides size and connectivity changes, MRAI recommendations have also evolved, with suggestions to reduce or eliminate it [4]. A survey among operators showed that 90% of respondents reported to have disabled MRAI [5]. Vendors as Juniper, Arista or Brocade are also disabling MRAI by default (although others keep the standard value).

In this letter we revisit the measure of route propagation times. We define metrics related with the advertisement of new routes, and metrics characterizing route withdrawal. We apply the defined metrics to the public dataset generated by RIPE BGP beacons and collectors. In this case, clock synchronization cannot be taken for granted, so we develop a model to provide an upper bound for the clock offset error for each of the measures taken. We describe the evolution of the route propagation delay between 2012 and 2018.

II. BEACONS AND COLLECTORS

RIPE RIS (Route Information Service) Beacons (*beacons*, for short) are BGP prefixes advertised (UP event) and withdrawn (DOWN event) at fixed times from different locations. Most of the sources advertise both an IPv4 and an IPv6 beacon. For each beacon, a corresponding *anchor* prefix is advertised; as the anchor prefix is not scheduled for any change, it allows detecting transient network events.

Until 2017 all RIS beacons were advertised by Quagga software routers with default configuration. In particular, MRAI was set to 30s. The Quagga version used featured a MRAI implementation compliant with the old BGP specification, RFC1771, in which route advertisements were delayed, but route withdrawals were not (in the current specification, MRAI is applied to both types of BGP messages [1]). MRAI in Quagga is implemented in a per-peer basis, meaning that a single timer per peer is applied to all prefixes. This means that route propagation to a peer is driven by the periodic expiration of the MRAI timer. Thus, the advertisement messages originated for an UP event can experience a delay between 0 and 30s, while the withdrawn messages that are sent to the direct neighbors in the DOWN event are not delayed.

In the period from November 2017 to March 2018, RIPE underwent changes to upgrade its infrastructure [6] that re-

sulted in the removal of MRAI for any message issued by the routers responsible for beacon origination. We conclude that prior these changes there was a systematic delay for UP events that is not associated to the network tested, but to the testing infrastructure itself, and DOWN events were free from this systematic interference. From this change, both UP and DOWN events are free from this effect.

Public BGP *collectors* (e.g., RIPE RIS and Routeviews) receive advertisements from BGP routers, called *monitors*. As the collectors peer to the monitor as to a regular router, the monitor provides its best route to the destination. Collectors add a timestamp value, with a second granularity, for each BGP update received, and dump the route information into a file. In addition, collectors periodically dump the whole routing state inferred from the messages received.

III. ROUTE PROPAGATION METRICS

We define two relevant delay metrics related with an UP event between a source and a destination router: the *prefix reachability interval* and the *preferred route interval*. The *prefix reachability interval* indicates the time since the UP event started until the first advertisement for the prefix is received, and thus a route to the destination exists. The *preferred route interval* accounts for the reception at the destination router of the most preferred route to the prefix among those activated by the event, since the time the UP event was triggered. After this, the router does not change its forwarding decision for this destination prefix, unless a new routing event occurs. In our measuring infrastructure, this is the time since the UP event started until the last advertisement message sent by the monitor is received at the collector. The monitor may receive more messages, but are not transmitted as they are less preferred.

For a DOWN event we define the *prefix withdrawn interval* metric. This is the time it takes the monitor to determine that there is no valid route to the destination. The prefix withdrawn interval is measured as the time since the DOWN event started until the withdrawn message that removes the last route to a destination is received. Although realizing that a non-working route is in fact removed may seem futile, as it seems there is not much to fix at the BGP layer, it is relevant in some configurations: When Traffic Engineering is performed by advertising prefixes that are more specific than a base one (and thus, more preferred according to the longest-prefix match rule), the base prefix is typically advertised as a backup route. Then, the time to recover from a failure affecting the more specific prefix depends on the time to completely withdraw it.

As the measures depend on the time registered at the collector traces minus the time expected for the activation of the event. Therefore, they are affected by the clock offset error.

IV. CLOCK SYNCHRONIZATION BETWEEN BEACONS AND COLLECTORS

Mao et al. [7] observed that the offset between the clock of the sender and the receiver of the BGP route may affect the measure. In particular, they reported negative measured delays, that indicated a clock synchronization problem. There is also more recent anecdotic evidence on clock synchronization

issues affecting to BGP route propagation measurement in [8]. Because of their findings regarding errors, Mao et al. did not attempt to measure the propagation time, and they focussed in *convergence time*, i.e., the interval between the first and the last message for an event. As both times are measured by the same clock, the value is consistent. Many other studies have used beacons and collectors to study BGP route exchange, reporting route convergence times but not propagation times ([9], [10], [11]). Convergence times are also reported for routes different than beacons (e.g., [12]). In this section we analyze the clock offset between beacon sources and BGP route collectors, to enable its use for route propagation measurement.

We assume that beacons are originated in the same nodes that annotate the time at which the BGP updates are received, as occurs for the RIS infrastructure. Nodes C_i and C_j send and withdraw beacons at regular intervals, and each one receives the beacon of the other from its corresponding monitors. Each beacon event is denoted as e . Collector C_j receives an update from a monitor for a beacon event e from source C_i through path p , and registers it with time $T_{i,j,p}^e$, measured by the local clock at C_j . This update can be either an advertisement or a withdrawn message related with the event. The event started at time T_i^e , measured with the local clock of C_i . Lets assume there is an offset $\delta_{i,j}^e$ between the clocks of C_i and C_j at the time this beacon event e started. $\delta_{i,j}^e$ is positive if the clock of C_j shows a larger value than C_i , negative otherwise. $D_{i,j,p}^e$ is the propagation delay for the transmission of an update for a beacon event e from C_i to C_j through path p . Then, $T_{i,j,p}^e = T_i^e + \delta_{i,j}^e + D_{i,j,p}^e$. Note that $\delta_{i,j}^e$ does not depend on the path followed by the update. With the data available at collector C_j and the information about the schedule for beacon advertisements (defined at the local time of the beacon source C_i) we can measure and compute different values of $T_{i,j,p}^e - T_i^e$, therefore obtaining different measures for $\delta_{i,j}^e + D_{i,j,p}^e$.

Lets consider the opposite direction, C_i receiving a beacon from C_j . In this case $T_{j,i,p'}^e - T_j^e = -\delta_{i,j}^e + D_{j,i,p'}^e$. Lets take the minimum time, the time for the first message received related with beacon event, over the many paths for a beacon event e , $m_{i,j}^e = \delta_{i,j}^e + \min_p (D_{i,j,p}^e)$ and its corresponding value on the other collector, $m_{j,i}^e = -\delta_{i,j}^e + \min_{p'} D_{j,i,p'}^e$. As the contribution of the term $\delta_{i,j}^e$ must be positive in one of the expressions, and propagation delays are positive values, we can set the following upper bound for the absolute value of the clock offset error:

$$|\overline{\delta_{i,j}^e}| = \max(m_{i,j}^e, m_{j,i}^e) \quad (1)$$

This bound depends on the propagation time of the fastest message from C_i to C_j , and from C_j to C_i . If many paths exist between C_i and C_j , it could be expected that at least one of the paths is fast enough to generate a low value for this bound. However, if a high propagation delay exists between two nodes i and j , at least in one direction, the value for $|\overline{\delta_{i,j}^e}|$ can be too pessimistic. In this case, a better bound results from taking into account that $\delta_{i,j}^e = \delta_{i,k}^e + \delta_{k,j}^e$, for a third collector/beacon source C_k , i.e., a *triangular* bound. Therefore, the following expression may provide a better estimation

$$|\overline{\delta_{i,j}^e}| = \min_k (|\overline{\delta_{i,k}^e}| + |\overline{\delta_{k,j}^e}|) \quad (2)$$

V. ANALYSIS OF BGP BEACON TRACES

The code used for this section is available at <https://github.com/albertogarciam/BGPbeacon-convergence/>. This code uses RIS collector data publicly available.

A. Description of the dataset

We analyse the data traces for beacon prefix advertisements provided by RIS collectors in the first 30 days of October for different years. As RIS collectors also generate beacon advertisements, we can apply to this dataset the clock offset error estimation methodology presented in the previous section. We select a subset of collectors, and their corresponding beacons, that have been reported to be active for the last 10 years. The list of collectors is RRC00, RRC01, RRC04, RRC05, RRC06, RRC07, RRC10, RRC11, RRC12, RRC13, RRC14 and RRC15. We excluded RRC03, as its beacons had a different activation period for some years, and RRC16, which was inactive between 2012 and 2016.

RIS collectors dump routing tables daily at 00:00, 08:00 and 16:00. The inspection of the files containing the updates indicates large gaps without any timestamp annotation for the time during which the dump process lasts. For example, for collector RRC00, the first update for the file of 1st Oct. 2009, 00:00, appears at 00:18, i.e., 18s after the start of period associated to the file. This large gap is unusual in the traces of the collector. This time corresponds to the beacon advertisement scheduling, so that no beacon can be registered before this time. Fortunately, some beacon events are scheduled when RIS collectors are not dumping the routing table, i.e., at 02:00, 04:00, 06:00, 10:00, etc. After the RIS infrastructure upgrade commented before (see also [6]), route recording is no longer affected by RIS BGP dumps. To obtain an homogeneous set of measures free from interference for all the years considered, we select UP events triggered at 4:00, 12:00 and 20:00, and the DOWN events triggered 6:00, 16:00 and 22:00 (90 events of each type for each measurement period). The beacon traces are sanitized by skipping all data from the periods in which the monitor detected activity for the corresponding anchor. We also remove the data for a monitor/beacon period if there was anchor activity up to 10 minutes before the start of the event.

B. Clock offset estimation

We apply the methodology proposed in Section IV to the dataset to compute $|\delta_{i,j}^e|$ for every pair of collectors and for each event. For this analysis, we only consider DOWN events, as UP events before 2017 are inflated by the MRAI at the beacon source. Figure 1 shows the 90th-percentile of the estimations over the measured events for each collector pair, ordered in increasing values. With the 90th-percentile we aim to represent a close to worst-case scenario for the clock offset error while removing the outliers. We exclude the pairs formed by the same collector (the collector storing info about the beacon it has generated), as there is no clock error in this case. We depict the value computed for Oct., years from 2011 to 2018; note some collector/beacons have experienced issues over the time, so that there is no data for all the collector

pairs. For example, in 2012 RRC06 did not collect data, the beacons for RRC04 were not advertised, and the anchors of RRC07 were advertised at the same time as the beacons, invalidating the data.

We observed that the fraction of error bounds improved by the triangular bound (Equation 2) ranges from 0.13 (2011) to 0.04 (2016).

C. Route propagation measurement

We measure route propagation for 2012 and 2018, as these years exhibit relatively low values for the 90th-percentile of the clock offset error bound. Only monitor/beacon pairs with more than 45 observations of UP and more than 45 DOWN events (more than half of the possible observations) are considered. This results in 752 pairs for IPv4 beacons and 482 for IPv6 in 2012, 620 and 588 respectively for 2018.

We filter out *zombie* routes and advertisements experiencing RFD. We consider that routes remaining for more than 90 mins after a DOWN event are *zombies*, according to the criteria of Fontugne et al. [13]. Zombie routes are caused by routers not propagating the changes, at least for this duration, probably due to router malfunction. BGP messages arriving after 20 mins may have suffered RFD (default values result in a minimum route suppression time of 30 min for Cisco and 21 min for Juniper). Unstable ASes causing periodic BGP updates may also result in updates in the period between 20 min after the current event was triggered and the start of the next event, according to the frequencies reported by Al-Musawi et al. [14]. For the rest of the analysis we remove the whole monitor/prefix data for events with updates arriving after 20 mins, eliminating both RFD and zombie events. Table I shows a summary of the results related with zombie prefixes and RFD. While the incidence of zombie routes has been reduced (dramatically for IPv4), updates arriving after 20 mins are now more frequent, with more monitor/prefix pairs affected.

Table II shows the mean for the 50th-percentile of the measures for the metrics defined in Section III. Apparently, propagation times related with UP events have improved over the years, but we have to consider that the generation of UP events was affected by an additional MRAI in 2012, not related with the network measured, but with the measuring infrastructure. The 50th-percentile of a random delay

Fig. 1: 90th-percentile of the upper bound for the clock offset error.

	IPv4 2012	IPv6 2012	IPv4 2018	IPv6 2018
<i>Zombie prefix</i> events	0.6%	1.0%	0.03%	0.5%
Possible RFD events	0.3%	0.1%	1.3%	3.4%
Monitor/beacon pairs observing at least one possible RFD event	34%	25%	89%	57%

TABLE I: Percentage of *zombie prefix* events in DOWN events, possible RFD occurrences in any event and monitor/prefix pairs observing at least one possible RFD event.

uniformly distributed between 0 and 30 is 15, so a first (rough) approximation would be to consider that the 50th-percentile propagation measures of 2012 are increased by this 15s value. For the withdrawn metric, which is directly comparable in both measuring periods, we observe an increase over time for IPv4 and a decrease for IPv6. For 2018, IPv6 route propagation is much faster than IPv4's in the three metrics considered.

We now take a closer look on the results. Figure 2 depicts the CDF of the prefix reachability interval for each monitor/prefix pair. For the IPv4 case, we also depict the maximum and minimum values according to the 90th-percentile of clock offset error for each pair (the IPv6 case is similar). The mean value of the 90th-percentile of the error for the represented collector/beacon pairs is relatively low compared to the value of the measured times (3.9s for IPv4, 2012; 2.4s for IPv4, 2018). Data for 2012 and 2018 are roughly similar, after considering the increase resulting from the MRAI delay for 2012 (up to 15s) incurred by the beacon generator. There is a small fraction of pairs with high values for the 2018 case. For IPv4, a single monitor, out of a total of 54, experiences high delays (more than 75s) for all beacons observed. An IPv6 monitor of the same AS to the IPv4 one is also responsible for most of the large propagation delays for IPv6.

For the prefix reachability interval (Figure 2), we can identify some performance classes. For 2018, IPv4, half of the values are close to 0. Then, a second plateau appears, 15s higher, which may result from an AS applying an MRAI delay of 30s for the first path arriving (see the comment made before for the effect of the MRAI in a 50th-percentile measure). We observe a lower fraction of IPv6 monitor/beacon pairs compared to IPv4 whose fastest path is affected by MRAI.

For the rest of the figures, we skip the information about the clock error bound: as the absolute values measured increase, the relative impact of this error is lower.

We show the results for the preferred route interval in Figure 3. The influence of MRAI is less clear than for the route propagation case, but for 2018 data there is a visible gap between few seconds (no MRAI) and 15s, and then a short plateau around 30s.

	IPv4 2012	IPv6 2012	IPv4 2018	IPv6 2018
Prefix reachability interval	22.9	20.7	12.0	8.8
Preferred route interval	34.2	27.5	20.4	16.17
Prefix withdrawn interval	90.3	111.6	115.4	94.3

TABLE II: Mean values for the 50th-percentile of prefix reachability, preferred route and prefix withdrawn intervals, measured in seconds.

Fig. 2: CDF of the prefix reachability interval for IPv4 with clock error bound, and for IPv6.

Fig. 3: CDF of the preferred route interval.

Finally, the prefix withdrawn interval is depicted in Figure 4. In this last figure we have removed the contribution of two monitors, with intervals over 600s for all beacon prefixes observed (both IPv4 and IPv6). These monitors also contributed with high intervals in the prefix reachability and preferred route interval figures. A small fraction of the prefixes (below 2% in all cases) are withdrawn in less than 30s, and also few routes in less than 60s (below 8% for IPv4, 2018). This may suggest that at least a path experiencing MRAI is found for most of the beacon/collector pairs considered (remember that BGP needs to flush the routes from all possible paths until the route is eventually withdrawn). The figure also indicates that IPv6 performs better than IPv4 for 2018, with a systematic improvement of around 30s. Besides, IPv4 has worsened its performance from 2012 to 2018, roughly in the same amount.

We now summarize the results. From 2012 to 2018, we observe that mean values of the prefix reachability and preferred route intervals are similar once removed the effect of the MRAI in the ASes injecting the beacons for 2012. For these metrics, the graphs show that a larger fraction of paths seem to traverse less ASes that apply MRAI. However, regarding to the prefix withdrawn interval, we observe that it improves for IPv6, but it worsens for IPv4. If we compare protocols for the same year, IPv6 route propagation outperforms IPv4.

Finally, we hypothesize about possible causes for the differences observed between years and protocols. We identify two major driving forces for changes in the route propa-

gation performance: path diversity and MRAI configuration. Regarding to path diversity, as the number of paths between a beacon source and a collector grows, the chance that a slower path appears increases, increasing the prefix withdrawn interval, as a route is not withdrawn until the last possible path is removed. Regarding to MRAI configuration, the expected effect of reducing the number of ASes applying it would be a reduction in all measured intervals.

The increase in the prefix withdrawn interval for IPv4 from 2012 to 2018 suggests an increase in the path diversity. There are some evidences supporting this claim: we observe that the mean number of ASes observed during a DOWN event has increased from 6.9 (2012) to 8.1 (2018); the mean number of paths explored when removing a route has increased from 2.7 to 3.5; and the number of different intermediate ASes for the last route withdrawn has increased from 3.3 to 3.6. On the other hand, path diversity for IPv6 has reduced in the same period of time (e.g., the last route withdrawn traversed 4.5 ASes in 2012, and 3.4 in 2018) for the source and destination peers observed, and this is compatible with the reduction in the route withdrawn interval measured. Note that even if the fraction of ASes configured with MRAI may have decreased over time for IPv4, the net effect in the prefix withdrawn interval could be compensated by the the variation in path diversity. Further study is required to separately evaluate path diversity and MRAI configuration, and to determine why IPv4 and IPv6 behave differently in the data analysed.

VI. CONCLUSIONS

We presented a methodology to measure the route propagation time with BGP route collectors and beacons, which includes the influence of clock offset mismatches between route source and destination. We also took into account some issues that affect the quality of the BGP traces related with beacons and collectors. In general, the upper bound for the clock offset error between beacon sources and collectors has been reduced over the years, as a consequence from an improvement of the best case of the route propagation time between collectors.

We analyzed the propagation time for Oct. 2012 and 2018, both periods with low clock offset error values, to conclude that the prefix reachability and preferred route intervals have not varied much (once accounted for the difference in the

beacon generation process). However, the prefix withdrawn interval has increased for IPv4. Route propagation largely depends on both path diversity and MRAI configuration.

More work needs to be done on interdomain routing propagation and convergence, as few changes have occurred in the last years and there still is room for improvement, according to the measures presented.

REFERENCES

- [1] Y. Rekhter, S. Hares, and T. Li, "A Border Gateway Protocol 4 (BGP-4)," RFC 4271, Jan. 2006.
- [2] C. Villamizar, R. Chandra, and D. R. Govindan, "BGP Route Flap Damping," RFC 2439, Nov. 1998.
- [3] C. Labovitz, A. Ahuja, R. Wattenhofer, and S. Venkataschary, "The impact of internet policy and topology on delayed routing convergence," in *IEEE INFOCOM*, 2001, pp. 537–546.
- [4] A. Fabrikant, U. Syed, and J. Rexford, "There's something about MRAI: Timing diversity can exponentially worsen BGP convergence," in *IEEE INFOCOM*, 2011, pp. 2975–2983.
- [5] P. Gill, M. Schapira, and S. Goldberg, "A survey of interdomain routing policies," *Computer Communication Review*, vol. 44, no. 1, pp. 28–34, 2014.
- [6] C. Petrie. (2015) Updates to the RIPE NCC Routing Information Service. [Online]. Available: https://labs.ripe.net/Members/colin_petrie/updates-to-the-ripe-ncc-routing-information-service
- [7] Z. M. Mao, R. Bush, T. G. Griffin, and M. Roughan, "BGP beacons," in *ACM SIGCOMM conference on Internet Measurement*, 2003, pp. 1–14.
- [8] Ben Cox. (2019, May) The speed of BGP network propagation. [Online]. Available: <https://blog.benjojo.co.uk/post/speed-of-bgp-network-propagation>
- [9] B. Zhang, D. Massey, and L. Zhang, "Destination reachability and BGP convergence time," in *IEEE GLOBECOM*, 2004., vol. 3, 2004, pp. 1383–1389.
- [10] D. Pei and J. Van der Merwe, "BGP convergence in virtual private networks," in *ACM SIGCOMM Conference on Internet Measurement*. ACM, 2006, pp. 283–288.
- [11] R. Oliveira, B. Zhang, D. Pei, and L. Zhang, "Quantifying path exploration in the internet," *IEEE/ACM Transactions on Networking (TON)*, vol. 17, no. 2, pp. 445–458, 2009.
- [12] G. Huston. (2019) BGP in 2018. Part 2 - BGP Churn. [Online]. Available: <http://www.potaroo.net/ispcol/2019-01/bgp2018-part2.html>
- [13] R. Fontugne, E. Bautista, C. Petrie, Y. Nomura, P. Abry, P. Gonçalves, K. Fukuda, and E. Aben, "BGP zombies: an analysis of beacons stuck routes," in *Int. Conf. on Passive and Active Network Measurement*. Springer, 2019, pp. 197–209.
- [14] B. Al-Musawi, P. Branch, and G. Armitage, "Recurrence behaviour of BGP traffic," in *IEEE Intl. Telecommunication Networks and Applications Conference (ITNAC)*, 2017, pp. 1–7.

Fig. 4: CDF of the prefix withdrawn interval.