Network Working Group
Internet-Draft
Intended status: Best Current Practice
Expires: January 14, 2021

M. McBride
Futurewei
D. Madory
Oracle
J. Tantsura
Apstra

July 13, 2020

AS-Path Prepend
draft-mcbride-grow-as-path-prepend-00

Abstract

AS_Path prepending provides a tool to manipulate the BGP AS_Path attribute through prepending multiple entries of an AS. AS_Path prepend is used to deprioritize a route or alternate path. By prepending the local ASN multiple times, ASes can make advertised AS paths appear artificially longer. Excessive AS_Path prepending has caused routing issues in the internet. This document provides guidance,to the internet community, with how best to utilize AS_Path prepend in order to avoid negatively affecting the internet.

Status of This Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at https://datatracker.ietf.org/drafts/current/.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on January 14, 2021.

Table of Contents

1. Introduction

The Border Gateway Protocol (BGP) [RFC4271] specifies the AS_Path attribute which enumerates the ASs that must be traversed to reach the networks listed in the BGP UPDATE message. If the UPDATE message is propagated over an external link, then the local AS number is prepended to the AS_PATH attribute, and the NEXT_HOP attribute is updated with an IP address of the router that should be used as a next hop to the network. If the UPDATE message is propagated over an internal link, then the AS_PATH attribute and the NEXT_HOP attribute are passed unmodified.

A common practice among operators is to prepend multiple entries of an AS (known as AS_Path prepend) in order to deprioritize a route or a path. This has worked well in practice but the practice is increasing, with both IPv4 and IPv6, and there are inherit risks to the global internet especially with excessive AS_Path prepending. Prepending is frequently employed in an excessive manner such that it renders routes vulnerable to disruption or misdirection. AS_Path prepending is discussed in Use of BGP Large Communities [RFC8195] and this document provides additional, and specific, guidance to operators on how to be a good internet citizen with the proper use of AS_Path prepend.

1.1. Requirements Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC 2119 [RFC2119].

2. Problems

Since it it so commonly used, what is the problem with the excessive use of AS_Path prepend? Here are a few examples:

2.1. Excessive Prepending

The risk of excessive use of AS_Path prepend can be illustrated with real-world examples. Consider the Ukranian prefix (95.47.142.0/23) which is normally announced with an inordinate amount of prepending. A recent analysis revealed that 95.47.142.0/23 is announced to the world along the following AS path:

3255 197158

In this example, the origin AS197158 appears 23 consecutive times before being passed on to a single upstream (AS3255), which passes it on to the global internet, prepended-to-all. An attacker wanting to intercept or manipulate traffic to this prefix might enlist a datacenter of questionable morals who would allow announcements of the same prefix with a fabricated AS path such as 999999 3255 197158. Here the fictional AS999999 represents the shady datacenter. This malicious route would be pretty popular due to the shortened AS path length and might go unnoticed by the true origin, even if route-monitoring had been implemented. Standard BGP route monitoring checks a route's origin and upstream and both would be intact in this scenario. The length of the prepending gives the attacker room to craft an AS path that would appear plausible to the casual observer, comply with origin validation mechanisms, and not be detected by off-the-shelf route monitoring.

2.2. Prepending during a routing leak

In April 2010, China Telecom experienced a routing leak. While analyzing the leak something peculiar was noticed. When we ranked the approximately 50,000 prefixes involved in the leak based on how many ASes accepted the leaked routes, most of the impact was constrained to Chinese routes. However, two of the top five most-propagated leaked routes (listed in the table below) were US routes. Was there some grand conspiracy to intercept traffic destined for

these routes? Actually, it was due to something much more troubling: gratuitous AS path prepending.

During the routing leak, nearly all of the ASes of the internet preferred the Chinese leaked routes for 12.5.48.0/21 and 12.4.196.0/22 because, at the time, these two US prefixes were being announced to the entire internet along the following excessively prepended AS path: 3257 7795 12163 12163 12163 12163 12163 12163. With this odd configuration, virtually any illegitimate route, whether a deliberate hijack or an inadvertent leak, would be preferred over the legitimate route. In this case, the victim is all but ensuring their victimhood.

There was only a single upstream seen in the prepending example from above, so the prepending was achieving nothing while incurring risk of hijacked traffic during a routing leak or hijack. You'd think such mistakes would be relatively rare, especially now, 10 years later. As it turns out, there is quite a lot of prepending-to-all going on right now and during leaks, it doesn't go well for those who make this mistake. While one can debate the merits of prepending to a subset of multiple transit providers, it is difficult to see the utility in prepending to every provider. In this configuration, the prepending is no longer shaping route propagation. It is simply incentivizing ASes to choose another origin if one were to suddenly appear whether by mistake or otherwise.

2.3. Route Competition

So what happens when a non-prepended route competes against an excessively prepended route? Let's consider a real-world example. The Polish route 91.149.240.0/22 is normally announced with the origin prepended three times (41952 41952 41952) to three providers and prepended twice to a fourth. Beginning at 15:28:14 UTC on June 6, a new origin that was not prepended appeared in the routing table for this route. As is illustrated in the graphic below, AS60781 quickly became the most popular version of this route for the next week until it disappeared.

When both AS41952 and AS60781 were in contention for being considered the origin of this prefix, the non-prepended route was dominating as we would expect. In some cases, the impact of prepending isn't as straightforward. Let's take 66.220.224.0/19 as an example. This prefix is prepended but isn't one of the 60,000 prepended-to-all routes mentioned earlier because its prepending is only visible to a little more than half of our BGP sources. In any event, this prefix is announced to the internet in two ways: it's prepended to AS6939 and not prepended to AS174:

...6939 17356 17356 17356 17356 17356 17356 17356 17356 17356 17356 17356

...174 17356

From these two route options, one might reasonably infer that it is 17356's intention to deprioritize routes to AS6939 by prepending itself 10 times on routes to that upstream. It may seem to follow that the non-prepended path to AS174 would be the most popular. However, the opposite is true. Despite extensive prepending, AS6939 is the more popular choice. In this case, prepending is going up against the local preferences of a legion of ASes: AS6939 has an extensive peering base of thousands of ASes. These ASes opt to send traffic for free through their AS6939 peering links instead of having to pay to send traffic through a transit provider (and via AS174) regardless of the AS path length. AS17356 could prepend their routes to AS6939 100 times (please don't!) and AS6939 would still be the more popular provider. Keep in mind that the average AS diameter of the internet is only around 4 hops, so prepending more than a couple of times buys you nothing.

2.4. Prepending to All

Out of approximately 750,000 routes in the IPv4 global routing table, nearly 60,000 BGP routes are prepended to 95% or more of hundreds of BGP sources. About 8% of the global routing table, or 1 out of every 12 BGP routes, is configured with prepends to virtually the entire internet. The 60,000 routes include entities of every stripe: governments, financial institutions, even important parts of internet infrastructure.

Much of the worst propagation of leaked routes during big leak events have been due to routes being prepended-to-all. AS4671 leak of April 2014 (>320,000 prefixes) was prepended-to-all. And the AS4788 leak of June 2015 (>260,000 prefixes) was also prepended-to-all. Prepended-to-all prefixes are those seen as prepended by all (or nearly all) of the ASes of the internet. In this configuration, prepending is no longer shaping route propagation but is simply incentivizing ASes to choose another origin if one were to suddenly appear whether by mistake or otherwise. The percentage of the IPv4 table that is prepended-to-all is growing at 0.5% per year. The IPv6 table is growing slower at 0.2% per year. The reasons for using prepend-to-all appears to be due to 1) the AS forgetting to remove the prepending for one of its transit providers when it is no longer needed and 2) the AS attempting to de-prioritize traffic from transit providers over settlement-free peers and 3) there are simply a lot of errors in BGP routing. Consider the prepended AS path of 181.191.170.0/24 below:

52981 267429 267429 267492 267492 267429 267429 267492 267492 267429
267429 267492 267492 267429

The prepending here involves a mix of two distinct ASNs (267429 and 267492) with the last two digits transposed.

2.5. Memory

Some BGP implementations have had memory corruption/fragmentation problems with long AS_PATHS.

2.6. Errant announcement

There was an Internet-wide outage caused by a single errant routing announcement. In this incident, AS47868 announced its one prefix with an extremely long AS path. Someone entered their ASN instead of the prepend count 47868 modulo 256 = 252 prepends and when a path lengths exceeded 255, routers crashed

3. Best Practices

Many of the best practices, or lack thereof, can be illustrated from the preceeding examples. Here's a summary of the best current practices of using AS-Path prepend:

- Network operators should ensure prepending is absolutely necessary. Many of your networks have excessive prepending

- Prepending more than a couple of times buys you nothing. So don't do it.

- Prepending-to-all is a self-inflicted and needless risk that serves little purpose. Those excessively prepending their routes should consider this risk and adjust their routing configuration.

- It is not typical to see more than 20 ASes in a AS_PATH in the Internet today even with the use of AS_Path prepend. The Internet is typically around 5 ASes deep with the largest AS_PATH being 16-20 ASNs. Some have added 100 or more AS_Path prepends and operators should therefore consider limiting the maximum AS-path length being accepted

4. IANA Considerations

5. Security Considerations

There are no security issues introduced by this draft.

6. Acknowledgement

7. Normative References

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <https://www.rfc-editor.org/info/rfc2119>.

[RFC4271] Rekhter, Y., Ed., Li, T., Ed., and S. Hares, Ed., "A Border Gateway Protocol 4 (BGP-4)", RFC 4271, DOI 10.17487/RFC4271, January 2006, <https://www.rfc-editor.org/info/rfc4271>.

[RFC8195] Snijders, J., Heasley, J., and M. Schmidt, "Use of BGP Large Communities", RFC 8195, DOI 10.17487/RFC8195, June 2017, <https://www.rfc-editor.org/info/rfc8195>.

Authors' Addresses

Mike McBride
Futurewei

Email: michael.mcbride@futurewei.com

Doug Madory
Oracle

Email: douglas.madory@oracle.com

Jeff Tantsura
Apstra

Email: jefftant.ietf@gmail.com