August 2023
Geoff Huston

IEPG at IETF 117

This is part of a personal commentary on the meetings at the July 2023 meeting of the Internet Engineering Task Force (IETF 117). If you want to know what was presented and the comments at the mic see the IETF 117 meeting archive[1].

The IEPG meets for a couple of hours before each IETF meeting. It's a somewhat eclectic collection of presentations, with some vague common thread of relevance to Internet operations. Here's a summary of my impression from these IEPG session presentations for IETF 117[2].

IPv6 Hop-by-Hop Extension Headers - Again!

The saga over the level of support for IPv6 Extension Headers in the Public Internet continues, but at the same time the level of interest in the topic is gently waning, to the obvious relief of some!

The initial reports from 2015, reported (in RFC 7872) that the observed packet drop probability for IPv6 packets with extension headers was measured at rates between 11% and 55% of cases. For the subset of cases where the Fragmentation Extension Header was used, the drop rate was measured between 28% and 55%. The general observation was not of much consequence, since little in the public network had a high level of dependence on the availability of various IPv6 Extension Headers, but the consequences of drop in Fragmentation Headers was of greater import. Even there, the consequence is of little importance in most cases, except for UDP and other datagram protocols. Any robust TCP implementations using IPv6 should take into account received ICMP6 Packet Too Big signals and adjust their outbound MSS value to avoid sending fragmented IPv6 packets. However, UDP can't do this.

What uses UDP transport? Well, the DNS for one, and QUIC for another. QUIC circumvents the fragmentation issue by imposing an upper bound on 1,200 octets on packets, circumventing the problem. What's left is the DNS. The solution here is also to simply avoid using large DNS packets, and the purpose of DNS Flag Day 2020[3] was to change DNS server behaviour so as not to send large DNS packets that may require fragmentation. It appears to have been more tractable to change the application behaviour that had been encountering IPv6 packet fragmentation than it was to clean out the IPv6 paths that blocked the forwarding of fragmented packets!

The story with the network treatment of other forms of extension headers, and notably Destination Headers (DST) and Hop-by-Hop headers (HBH) is similar, but without any transport protocol measures that could mitigate the issue. Destination headers of 64 bytes in length or less appear to have an overall success rate of delivery in the public Internet of around 70% of IPv6 users[4], which is not exactly brilliant, but not that bad either. On the other hand, the story with HBH extension headers has proved to be a

[1] https://datatracker.ietf.org/meeting/117/agenda
[2] https://www.iepg.org/2023-07-23-ietf117/index.html
[3] http://www.dnsflagday.net/2020/
[4] https://stats.labs.apnic.net/v6frag/XA?hc=XA&hl=1&hr=1&hs=1&hx=1&hd=1&w=1&d=2

measurement challenge. What we've found in examining this behaviour in more detail[5] is that many network service providers simply drop IPv6 HBH packets. Why would they do this? The issue here is that all IPv6 packets with HBH settings need to be passed to the router's processor in every router that processes such a packet. In the context of the public Internet, this exposes the routing environment to a DOS attack where a high-volume packet flow with HBH headers has the capacity to dominate the router's packet processing queue thereby denying this processing capability to other packets and risks saturating the router's processing capability. Furthermore, the entire purpose of the HBH header provides a tool for one network to influence, or even override, the default forwarding and routing polices of another. Most network operators would see this as a less than desirable capability to pass on to edge connected hosts. Given that we've been unable to find a compelling use case for HBH packets in the public IPv6 Internet, then allowing routers to process such packets represents a risk without any clearly identified benefit. Little wonder that most networks in the public Internet take the prudent course of action and drop such packets.

The overall conclusion is that all extension headers in IPv6 in public Internet are just not sufficiently broadly supported to be useful, including fragmentation. Prudent advice is to simply stay away from this IPv6 "feature," if at all possible.

RPKI

These days IEPG meetings appear to include a mandatory session on IPv6 Extension Headers and a second presentation on some aspect of inter-domain routing, normally in the area of securing BGP. This IETF 117 IEPG session heard one individual's perspective on the status and future prospects on the deployment of a secure routing framework. Such presentations have ranged from the deeply pessimistic[6] () to the wildly optimistic, as was the case with this presentation from Job Snijders. This topic has a rich history going back to the start of inter-domain routing and the faltering progress over the decades has been variously attributed to a lack of focus and effort, or to the intractably complex nature of the problem in the first place, or some combination of the two!

It's a complex problem with many facets, including the operation of a credential scheme to allow public keys to be associated with the holder of the instruments of routing, IP addresses and Autonomous Numbers, the generation of authorities that convey permissions to originate routes into the inter-domain routing space, and attestations that relate to inter-domain adjacencies that allow AS Path attributes of a route to be audited in some manner, and then an overall protocol validation framework that attempts to create a protocol correctness model. It's been thirty years in the making so far, and Job's optimistic estimate is a further twenty years! A more pessimistic estimate tends towards the prediction that this work will be abandoned at some point! (I'd like to say that the likely outcome is somewhere in the middle, but I find myself somewhat pessimistic on this particular topic, so personally I'm leaning to the "not going to happen this way" side!)

Much of the effort has been devoted to building and operating the infrastructure of prefix origination credentials, the so-called Route Origination Attestations (ROAs). It's a mixed story here. Many network operators have now published ROAS and the total count of such credentials has reached some 40% of the total prefix object count, which is an encouraging situation[7]. Getting networks to filter BGP route updates and remove updates where the ROA credentials flag these routes as invalid is still work in progress. Only some 25% of Internet users are located behind such filtering networks, which indicates that much is still to be done. Of course, origination without some form of AS Path protection provides little in the way of protection against deliberate attack, and the work on AS Path protection is still a developing story with little to report in terms of deployed outcomes. The original effort, a comprehensive framework of linked signatures which is formed as an update propagates across the inter-AS space,

[5] https://www.potaroo.net/ispcol/2023-06/eh.html

[6] https://icann.zoom.us/rec/share/n073rIsK-5A6cjovaILbX8BCdySBOYtU4X65HmO5l9qGnn7brBhjiIblmDyymCKN.9q32Mcv2tsPLlZgd?startTime=1668609005000

[7] https://stats.labs.apnic.net/roa/XA?o=cXAl1r1v6trdpxv&t=Route+Objects&x=Valid&v=Total&d=Percent

termed BGPSEC, has not been deployed in any meaningful way, and the overheads associated with that approach tend to indicate that this picture is not going to change in the foreseeable future. An alternate approach using AS adjacency attestations filter by customer/provider relationships, is yet to come out of the IETF standards mill, and even so the somewhat reduced level of assurance on AS Path correctness is what caused the 20-year-old predecessor so-BGP model to founder, and this time around the level of assurance that folk will pick this up and run with it is even lower!

So, what are we chasing down here? Is the problem one of malicious orchestrated interference and routing chaos at a disturbing scale and frequency? Not so far. The dominant form of routing anomalies are simple route leaks. Instances of inadvertent operator configuration error that cause the local BGP configuration to advertise reachability in unintended ways. One argument goes along the line that if reducing the incidence of route leaks is the only outcome of this work then that's a Good Thing! An opposing view is that if we have to go to these extraordinary lengths to achieve what a decent BGP audit framework could do at much lower cost, then we have got this all wrong! (Perhaps we could invoke the Mighty Power of generative prediction models and apply Chat GPT to BGP!)

The entire area of securing routing suffers from the defender's dilemma. A defender has to defend everything, while the attacker need only concentrate their efforts on one point. Much of the model of BGP assumes that the actors who play BGP are all trustable, and attacks are "outsiders" who generate synthetic information to revert the operation of BGP. As Job described in a recent case, the attacker needs only to shift their locus of attack into one of impersonation and assume not only the identity of a "legitimate" network, but also assume control of all or part of the victim's network infrastructure. At this point the attacker is operating part of the network of the intended victim.

There is much in this space which is uncertain and even more which is only dimply understood. The continual stream of route leaks illustrates the observation that even without the overlay of security mechanisms the environment is sufficiently complex such that network operators just get it wrong from time to time. The desire to apply engineering solutions to this space is irresistible, but at the same time it can be quite unwise. Adding more steps and more interdependencies into an already complex distributed framework that experiences a volume of lapses caused by these operational complexities often just adds to the brittleness of the framework and does not do much at all to improve the overall robustness of the resultant system.

My takeaway from this is that securing the inter-domain routing system is one of those endless quests, alongside the quest for the fountain of youth and the holy grail! However, this is not necessarily bleak news. This lack of secured routing system has been with us from the outset, and we have accommodated it. We've learned not to trust the answer if all we've done is to send an IP packet to the "right" IP destination address. That's why we use HTTPS everywhere these days. If the remote part of a conversation can demonstrate that they have control of the name of the party we want to communicate with, and we encrypt the session, then the massive risks of an invisible "passing off" attack are mitigated. Today it does not matter how our encrypted packets reach their intended destination, as this aspect of path control of the communication is largely out of our control in any case. What matters is that we are communicating securely with the intended destination by virtue of their name, not by their address. This means that routing attacks are in essence just one more vulnerability in the highly populated world of denial-of-service attack vectors, and countering the entire class of such denial-of-service threats is a truly Sisyphean task!

BMP

As we've just observed the inter-domain routing space is deceptively complex. Observing that is going on in the routing environment, even at the level of an individual BGP speaker can extremely valuable. For many years we relied on implementation-specific "peepholes" that allowed a network operator to generate a feed of BGP events, or customisation of a BGP implementation to report on its internal state.

The IETF got around to standardising a BGP observation data feed, the BGP Monitoring Protocol (BMP) (RFC 7854).

This tool has been highly valuable, in that it does not only provide a feed of the BGP protocol transactions, but it can provide a feed of the current state of the BGP speak and its data structures. This presentation was a report on the implementation of improvements to the BMP implementation in the open source FRR BGP implementation.

In many ways this is not only the bread and butter of the IETF, but its entire reason for its existence. The standards produced by the IETF are intended to provide sufficient guidance to implementors to produce running code based on the specification, and also ensure that the code cleanly interoperates with other implementations of the same standard, Reports on such efforts to build code from the specification are a necessary and important aspect of the IETF's efforts.

Tweaking BGP

BGP is a member of a class of distributed peer flooding protocols, where the ultimate objective is to ensure that every member of the set of protocol speakers receives a copy of the protocol information irrespective of where the information was first injected into the set. Perhaps if this was all we wanted BGP would be a whole lot easier. But we have a penchant for adornment of technologies, and BGP is no exception.

In this case we are talking about the intentional limited flooding of BGP Path attributes, and the "misbehaviour" of BGP when it treats the attribute as a flooded attribute and pushes it well beyond its intended scope. This can cause an AS Path Attribute to be interpreted in the wrong context, such as black holes of misdirected packets.

One strategy to mitigate these sorts of issues can be explicit permit lists. Rather than simply propagate all unknown Path Attributes, lists of permitted Path Attributes by Attribute Type Code may be propagated through the network. This list may become bloated and inconsistent use may still lead to unintended consequences. Another possibility is to associate the feature to one or more listed ASes.

It's probably easy to bemoan the fact that we've got to this point in the adornment of BGP, but at the same time we need to recognise that BGP occupies a relatively unique role in the infrastructure of the Internet, and it's one of the few universally available signalling protocols that we have. We've got to here simply because there is no viable alternative!

Disclaimer

The above views do not necessarily represent the views or positions of the Asia Pacific Network Information Centre.

Author

Geoff Huston AM, B.Sc., M.Sc., is the Chief Scientist at APNIC, the Regional Internet Registry serving the Asia Pacific region.

www.potaroo.net