

October 2019  
Geoff Huston

## AS Prepending in BGP

In this article I'd like to look at one particular aspect of the Internet's inter-domain routing framework, namely the role of the Autonomous System (AS) Path in the operation of BGP, and in particular the use of AS Prepending.

### What is an Autonomous System?

To introduce this topic a brief introduction might be useful for those who are unfamiliar with this aspect of the routing infrastructure of the Internet

One of the most useful tools to allow a networked system to scale is hierarchies. Hierarchies allow the paraphernalia of detail within a bounded domain to be obscured outside of that domain. It's like looking at a set of maps of a country. At a national level the map would probably describe only the routes between cities, while a city map describes the layout of individual streets. The hierarchical abstraction in the Border Gateway Protocol (BGP) relies on the concept of the Autonomous System (AS). An AS encompasses a connected domain of connectivity where the paths between any two points that are located within the AS can be described by a path that itself lies entirely within that domain. The details of the internal topology within the AS are not used outside the AS. From the outside, an AS is just a set of address prefixes that summarise all the individual endpoints that are contained within that AS.

As well as a conceptual framework to describe hierarchies in the routing space, an Autonomous System also has a representation within BGP: it's just a simple number.

In a routing sense, an AS is an autonomous routing domain where an interior routing protocol is used to maintain the topology of interconnectivity within the AS. Each AS is free to select its own interior routing protocol. Larger networks may use IS-IS or OSPF, while smaller ASs may still use RIPv2 or EIGRP. Hybrid approaches are also common these days where the next hop addresses are managed by the IGP and all other addresses are maintained using iBGP.

With that, lets continue our focus on the inter-AS routing space occupied by eBGP.

## BGP and Route Selection

BGP is an example of a classic *Distance Vector* (DV) Protocol. In a DV protocol each speaker informs its adjacent neighbours of reachable address prefixes via a route advertisement and with each prefix it attaches a "cost". Each time a route is passed to a neighbour the cost is incremented. If a received advertisement offers a lower cost to an address prefix, then the router will select this path via the neighbour that informed the router of the lower cost path and pass on the address prefix and new cost to its neighbours.

It's a simple protocol, but it has a glaring problem. When connectivity to a prefix is withdrawn, the protocol may form a loop and endlessly pass the address prefix within the loop, incrementing the cost without limit. In Figure 1 when network B loses connectivity to network A then this does not mean that network B will drop its record of Network A. Network B has already announced that it can reach network A to networks C and D. When B loses its direct path to A (cost 1) it may learn a replacement path from network D (cost 4). As this is a new best path for B it will announce it to C, who will announce it to D, who will revise its announcement to B with cost 7, and so on. RIP used a simple but effective measure to stop this *count-to-infinity* by using a concept of a *maximum cost*. When the route metric reached this maximum cost, the route was discarded.

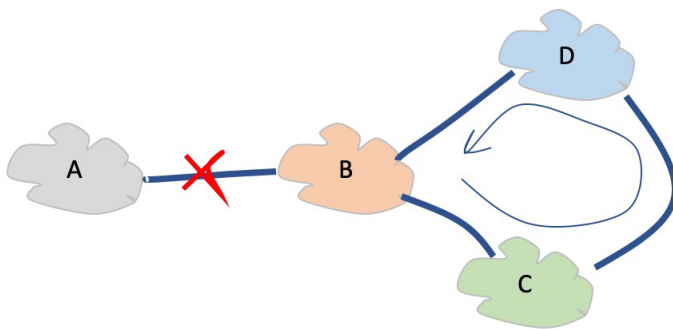


Figure 1 – Route Loop Formation

BGP uses a slightly different approach. Each time an eBGP speaker sends a route object to an adjacent AS it adds its own AS value to the AS Path attribute of the route object. This AS Path attribute has two uses in BGP. The first is in loop detection. If an AS sees its own AS in the AS Path, then it can conclude that this is a routing loop and immediately discard the route object.

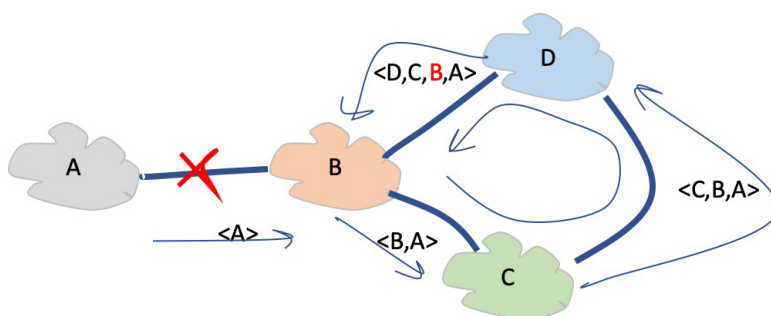


Figure 2 - Distance Vector with Path Tracing

The second use of the AS Path is as a comparative metric for route selection. The number of AS's in the AS Path, or the *Path Length*, acts as the route metric. Within a BGP route comparison process, when local policy preferences do not resolve the selection process then shorter AS Paths are preferred, as shown in Figure 3. In the case shown in this figure, AS E will receive route advertisements from both C and F. As the AS path length from C is one less than that from F, E will select the shorter path via AS C as the locally preferred path to reach prefixes advertised by A.

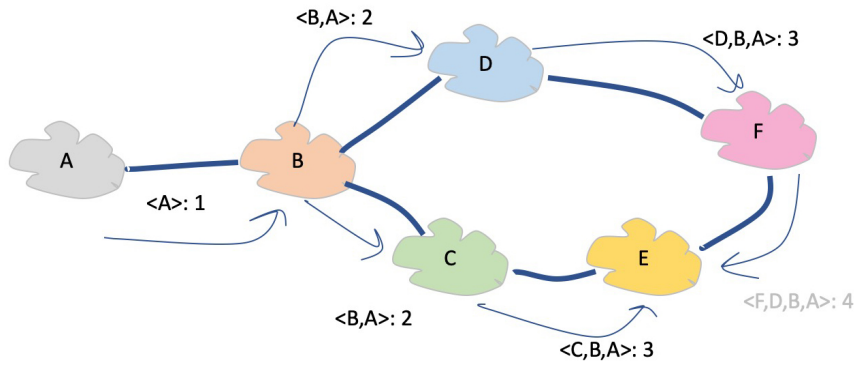


Figure 3 – AS Path Preference

A more complete enumeration of BGP's route preference algorithm is as follows (in order):

- Remove routes that list own AS in the AS Path.
- Remove routes where the next hop address cannot be resolved.
- Prefer the highest local-preference value.
- Prefer the shortest AS-path length.
- Prefer the lowest origin value.
- Prefer the lowest MED value.
- Prefer routes learned from an EBGP peer over an IBGP peer.
- Prefer best exit from AS.
- For EBGP-received routes, prefer the current active route.

This can be seen in practice in a snapshot from the routing table in one of the Route Views BGP collectors, shown in Figure 4. The route entry that contains an AS Path of length 3 is the selected “best” route. In this case anyone who was connected to this eBGP speaker would receive a single route advertisement for this prefix, containing the AS Path attribute <6447 1221 38803 56203>.

Network	Next Hop	Metric	LocPrf	Weight	Path
* 1.0.4.0/24	206.24.210.80	0	0	0	3561 209 4637 1221 38803 56203 i
* 1.0.4.0/24	212.66.96.126	0	0	0	20912 174 4637 1221 38803 56203 i
* 1.0.4.0/24	91.228.151.1	0	0	0	31019 6939 4637 1221 38803 56203 ?
* 1.0.4.0/24	85.114.0.217	0	0	0	8492 6939 4637 1221 38803 56203 e
* 1.0.4.0/24	94.156.252.18	0	0	0	34224 5580 4637 1221 38803 56203 i
* 1.0.4.0/24	154.11.98.225	0	0	0	852 6939 4637 1221 38803 56203 i
* 1.0.4.0/24	203.189.128.233	0	0	0	23673 6939 4637 1221 38803 56203 i
* 1.0.4.0/24	168.209.255.56	0	0	0	3741 6939 4637 1221 38803 56203 i
* 1.0.4.0/24	195.22.216.188	100	0	0	6762 4637 1221 38803 56203 i
* 1.0.4.0/24	103.247.3.45	0	0	0	58511 6939 4637 1221 38803 56203 i
* 1.0.4.0/24	173.205.57.234	0	0	0	53364 3257 4637 1221 38803 56203 i
* 1.0.4.0/24	95.85.0.2	0	0	0	200130 6939 4637 1221 38803 56203 i
* 1.0.4.0/24	162.243.188.2	0	0	0	393406 6939 4637 1221 38803 56203 i
* 1.0.4.0/24	66.185.128.1	162	0	0	1668 4637 1221 38803 56203 i
* 1.0.4.0/24	198.129.33.85	0	0	0	293 6939 4637 1221 38803 56203 i
* 1.0.4.0/24	5.101.110.2	0	0	0	202018 6939 4637 1221 38803 56203 i
* 1.0.4.0/24	67.17.82.114	2515	0	0	3549 3356 4637 1221 38803 56203 i
*> 1.0.4.0/24	<b>203.62.252.83</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>1221 38803 56203 i</b>

Figure 4 – Route Views Entry for an Address Prefix

There is also an address prefix selection algorithm that is used by the router's forwarding engine, to unconditionally prefer to use the route object that contains the most specific address prefix that encompasses the forwarded address, but we'll consider that later.

## Traffic Engineering via Prepending

BGP is first and foremost a self-learning topology maintenance protocol. Its objective is to maintain a set of forwarding tables that represent the “best” path to each address destination. “Best” refers to a

specific attribute in this context, namely that absent any other local policy settings, each path that BGP selects represents a path that transits the minimum number of AS networks to reach the destination. What BGP does not do is select the fastest path, or the one that has the greatest available capacity, the greatest level of stability, or one that represents the lowest financial cost to each network operator. All BGP does is pick what it believes is the shortest path in terms of the AS Path length.

How can a network operator alter BGP route advertisements to produce a different outcome?

One way is to selectively manipulate the routing along the less preferred paths by making the AS path longer in these less-preferred cases.

Figure 5 shows a simple topology where AS E is multi-homed with services from both C and F. B will normally prefer the path via C to send traffic to A, as this represents the shorter AS path for B.

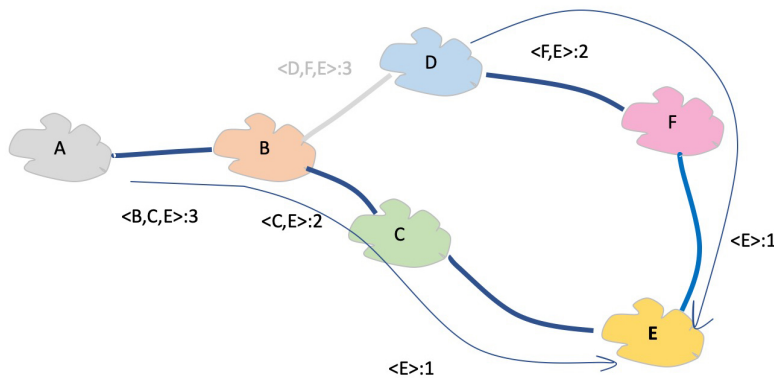


Figure 5 – A Multi-Homed AS

If E were to prepend a further two instances of its own AS number when advertising its routes to C, then B will now see a different situation, where the AS Path via D represents the shorter path (Figure 6). Through the use of selective prepending E is able to alter the routing decision of B, even though B is not an adjacent neighbour of E. The result is that traffic from A and B will be passed via D and F to reach E, rather than via C. In this way prepending implements *action at a distance* where the routing decisions made by non-adjacent AS's can be influenced by selective AS Path prepending.

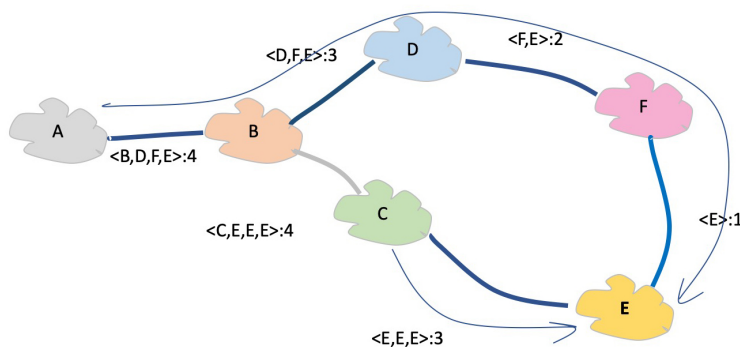


Figure 6 – AS Path Prepending

## Measuring Prepending in BGP

How common is AS prepending in the Internet?

If we look at a simple edge AS that is a non-transit stub AS, connected to a regional transit AS which itself is connected to a number of Tier-1 transit networks (a very common scenario in today's Internet) then the BGP route table had the following metrics for AS Paths:

Per-Prefix Paths	1,575,526
Unique Paths	217,585
Prepended Paths	52,774
Unique Paths with Prepending Stripped	204,573
Prefixes	803,040
Prefixes with Prepended Paths	198,328
Average Path Length	5.8
Average Path Length with Prepending Stripped	5.6
Average Prepend	2.8

Table 1 – Profile of AS Prepending Metrics for a BGP Stub AS

This profile indicates that AS prepending is relatively common, with some 25% of visible address prefixes carrying AS Paths that contain prepending, but the level of prepending appears to be relatively small.

The amount of prepending observed has the following distribution:

Prepend	Incidence	Relative	Cumulative
2	20,727	34%	34%
3	17,819	30%	64%
4	9,103	15%	79%
5	4,685	8%	87%
6	3,250	5%	92%
7	1,521	3%	95%
8	690	1%	96%
9	797	1%	97%
10	746	1%	98%
11	581	1%	99%
12	81	0%	99%
13	49	0%	99%
14	49	0%	100%
15	60	0%	100%
16	139	0%	100%
17	29	0%	100%
18	7	0%	100%
19	5	0%	100%
20	3	0%	100%
21	14	0%	100%
22	3	0%	100%
25	2	0%	100%
26	1	0%	100%
27	1	0%	100%
31	25	0%	100%

Table 2 – Profile of AS Prepending

One third of all prepending is a simple duplication of the AS, and a further third of all incidences of prepending sees the AS prepended twice. It is anomalous to see 25 instances of 31 prepended AS's and the utility of this level of prepending seems to be highly questionable.

Another perspective is to look at the distribution of AS Path Lengths. Figure 7 shows this distribution from the perspective of this stub AS.

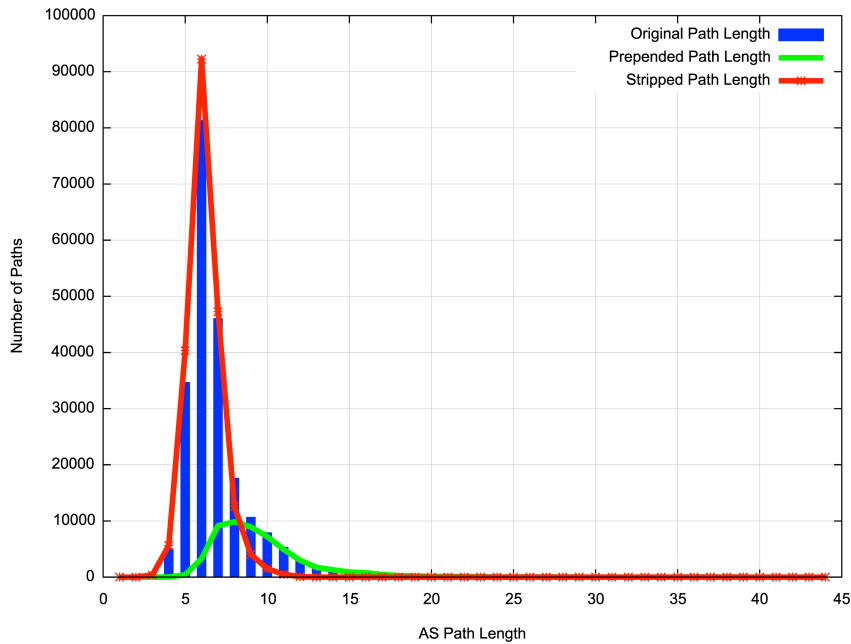


Figure 7 – AS Path Length Distribution

In this view of the routing table the majority of AS path are of length 6, while when looking at those paths that only contain prepending the paths of length 7 and 8 are the most common. If the prepending is stripped, the resultant path distribution is similar to the original distribution.

This data reflects the view of BGP from a single AS, 131072, located within APNIC Labs on the edge of the network. A slightly different view is obtained from looking at a Route Views BGP dump, as shown in Table 3.

Per-Prefix Paths	64,146,980
Unique Paths	4,052,146
Prepended Paths	1,008,096
Unique Paths with Prepending Stripped	3,797,600
Prefixes	832,076
Prefixes with Prepended Paths	676,841
Average Path Length	3.6
Average Path Length with Prepending Stripped	3.0
Average Prepend	2.8

Table 3 – Profile of AS Prepending Metrics for Route Views, AS6447

While the route collector has a significantly larger set of AS paths, as is expected, and a somewhat smaller average AS Path length, the prepending profile is largely similar, with an average level of prepending of 2.8.

The number of prefixes with prepended paths increases dramatically in this Route Views data set. A number of BGP peers of Route Views show a high level of path prepending, and what we see in the distribution of prepending is a significantly higher number of paths with just a single repeated AS in the AS Path. This aligns with our expectation, as prepending is intended to disfavour otherwise selected paths, allowing a non-prepended path to be preferred. It would be expected that in any single view of the routing system prepended routes would only be seen for “close” AS’s. When we look at large aggregated route sets, as it the case of Route Views, then the prepended path set is the union of what is “close” across the much larger set of Route Views BGP peers, so the level of visible prepending is expected to be larger than a single view.

Prepend	Incidence	Relative	Cumulative
2	263,363	52%	52%
3	88,671	17%	69%
4	81,754	16%	85%
5	26,339	5%	90%
6	20,356	4%	94%
7	9,362	2%	96%
8	3,794	1%	97%
9	5,050	1%	98%
10	4,750	1%	99%
11	4,243	1%	100%
12	228	0%	100%
13	286	0%	100%
14	318	0%	100%
15	186	0%	100%
16	470	0%	100%
17	164	0%	100%
18	22	0%	100%
19	15	0%	100%
20	10	0%	100%
21	33	0%	100%
22	60	0%	100%
23	17	0%	100%
24	3	0%	100%
27	9	0%	100%
31	100	0%	100%

Table 4 – Profile of AS Prepending in Route Views, AS6447

Of interest is that the distribution of larger prepend paths is largely similar in Route Views as is seen in a single view. Given that the effective diameter of the Internet is around 4 to 5 AS's, the incidence of large prepended AS Paths points to an observation that in many cases route selection is determined using local policy settings (such as preferring customer routes over peer or transit routes). The result is that prepending does not alter the path selection process in such cases and the prepended path remain the preferred path.

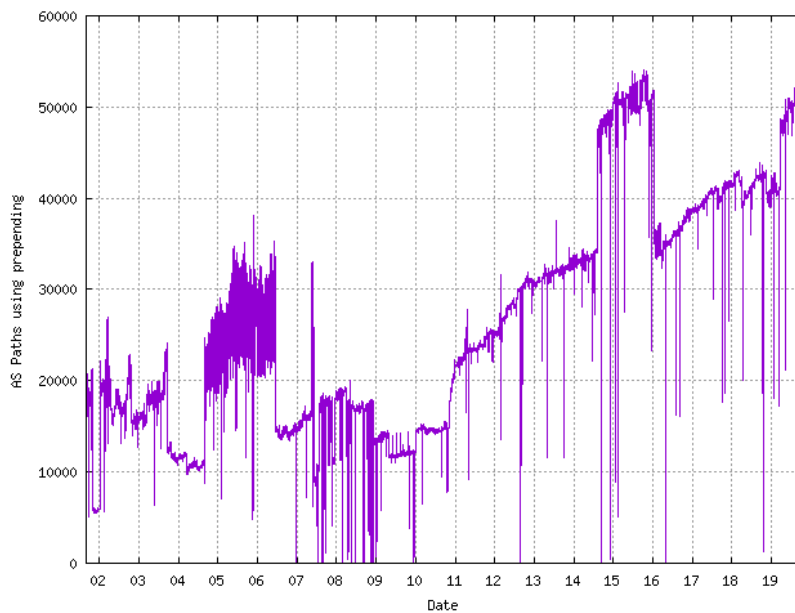


Figure 8 – Prepended AS Path Count as seen at AS131072 – 2001 – 2019



The profile of AS Path prepending has not been all that consistent over the past 20 years. In the period from 2001 to 2011 the level of visible prepending as seen at AS131072 remained within the band of 10,000 to 20,000 AS paths. This profile changed in 2011 and the number of paths containing prepending has risen over the ensuing nine years (Figure 8).

It appears that prepending is a relatively common traffic engineering technique in today's Internet.

However, it's useful to bear in mind that the purpose of AS Prepending is a localised mark of de-preferencing an alternate path that is adjacent to the prepending AS. The prepending is intended to cause more remote networks to prefer a path that does not contain the prepended AS's. The inference is that if we split a path into a leading section which is the AS path that occurred after prepending has taken place, then we would expect that the length of this leading section should be lower than the average stripped AS path length. In the case of the Stub AS with a single eBGP view the average length of the leading section of prepended AS Path is 2.3, while the average stripped AS path length is 5.6. This appears to confirm the expectation that AS prepending is more visible on routes that are originated close to the BGP observer, and prepending does not conventionally prevail globally when a route originator originates both prepended and non-prepended paths.

## Prepending Risks

Normally, AS Path prepending would not be expected to expose any heightened level of risk to the network.

However, when the network performs an excessive level of prepending then a "hostile" AS can set up unintended outcomes by performing local stripping of the prepended AS's in the AS Path and propagate the shortened AS path to its neighbours. The path appears perfectly plausible and the stripping of the prepending is invisible to the downstream AS, which makes it a challenge to observe. The result is potentially one of traffic re-direction.

From time to time there have been efforts to impose an operational consensus that sets a limit on either prepending or the total AS Path Length that is lower than the protocol-defined value of 256 maximum AS Path length. If we look at the maximum AS Path Lengths seen over time, then the incidence of prepending to bring the path Length to over 100 occurred only once between 1997 and 2008 but became commonplace from 2008 onward. A history of the maximum observed AS Path Length as seen at Route Views over this period is shown in Figure 9.

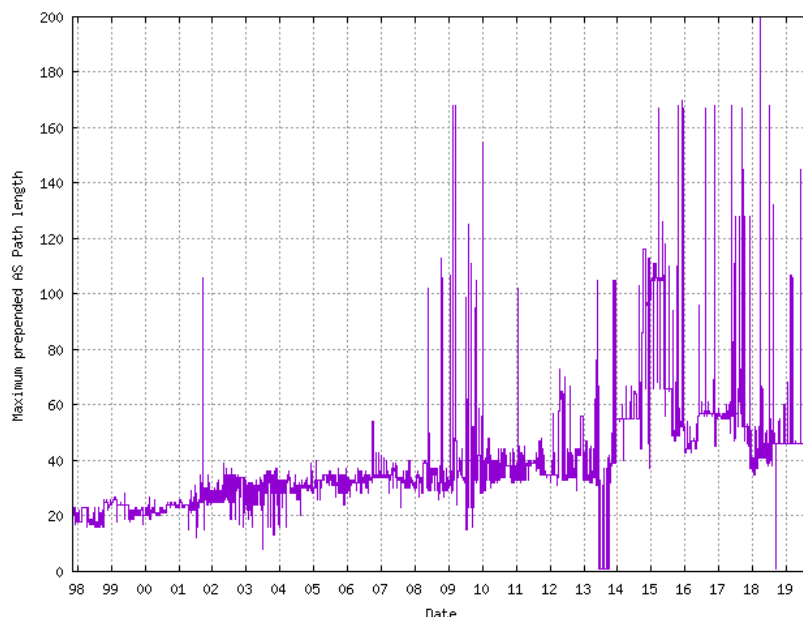


Figure 9 – Maximum AS Path Length observed by Route Views: 1998 – 2019



## Other Traffic Engineering Tools in BGP

AS Path prepending is perhaps a last resort in the traffic engineering toolbox, as other techniques operate in more predictable ways, while path prepending is a far more approximate approach.

Advertising more specifics to selected BGP neighbours has a more predictable outcome. In the BGP forwarding process more specific routes are always preferred over any covering aggregate prefix. A common approach is to advertise the covering aggregate address prefixes to all providers as a fallback, and then advertise more specific address prefixes to a subset of providers such that the incoming traffic meetings the locally determined policies (load balancing of incoming traffic, performance, minimising cost, and so on). More specific have occupied some 52% to 55% of the IPv4 routing table for the past two decades, while in IPv6 the count of more specific has risen to a current level of 47%.

A less common technique is *AS Path poisoning*. In this case rather than prepending the local AS to the prepended AS path (or in addition to conventional prepending) the AS adds the AS number of the network that it does not want to learn this route. Going back to the example in Figure 6, if network E did not want B to learn of a path via C to reach E, instead of prepending <E,E> it would prepend <B,E>, Network C would still learn and use such a route, but when it passed the route to network B with the path <C,E,B,E>, B would reject the route as a loop. AS Path poisoning is an uncompromising approach. Again, using the configuration in Figure 6, in a conventional prepending situation, if connectivity was disrupted with networks D or F, the prepended path via B would provide a backup path. However, with the poisoned path there is no backup provision, as B never learns of the alternate less preferred route via C.

BGP Communities provide a better approach to traffic engineering in BGP. A provider publishes a set of community values and associated route handling actions. If a customer adds a particular community value to the route that is advertised to the provider, the associated action is triggered within the provider network. A current example of community values and actions, taken from NTT's service offering, can be seen at <https://www.us.ntt.net/support/policy/routing.cfm>. Communities are not necessarily transitive and can only be counted on between adjacent AS's. Even so they can provide far better precision and far greater levels of control over traffic than can be achieved simply with more specifics and AS path manipulation.

It should be remembered that BGP has no feedback loop. Load balancing approaches that attempt to spread traffic across a number of end-to-end paths in order to evenly load the network and avoid hot spots are not supported by BGP. Load management protocols that integrate real time performance data into the management of traffic flows are used in some of the larger internal routing domains, particularly in large content feeder networks, but translating these dynamic capabilities and flow management algorithms into the realm of the inter-domain routing space is a particularly forbidding challenge in both administrative and protocol terms. The available tools we have for traffic engineering are pretty crude, and AS Prepending is no exception to that general characterisation. Precisely how much prepending is enough to achieve the wanted outcomes and how much is too much is a pretty much a shot in the dark.

It's hard to say whether routing administrators use prepending because it is observed to work as intended, or because they simply hope that it might work as intended!

---

## Disclaimer

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

---

## Author

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

*[www.potaroo.net](http://www.potaroo.net)*