Network Working Group
Internet-Draft
Intended status: Informational
Expires: March 29, 2013

Camilo Cardona
Pierre Francois
IMDEA Networks
September 25, 2012

Making BGP filtering an habit: Impact on policies
draft-cardona-filtering-threats-00

Abstract

This draft describes potential threats to the Internet routing policies of an autonomous system due to filtering of more specific BGP prefixes by its neighboring domains.

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 http://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 March 29, 2013.

Table of Contents

1. Introduction

It is common practice for network operators to propagate overlapping prefixes along with the prefixes that they originate. On the other hand, it can be beneficial for some Autonomous Systems (ASes) to filter overlapping prefixes (such operation needs to be translated into various requirements in order to be automatically performed) DRAFT-WHITE [1].

BGP makes independent, policy driven decisions for the selection of the best path to be used for a given IP prefix. However, in the data plane, the longest prefix match forwarding rule "precedes" the application of such policies. The existence of a prefix p' that is more specific than a prefix p in the Routing Information Base (RIB) will indeed let packets whose destination matches p' be forwarded according to the next hop selected as best for p' (the overlapping prefix). This process takes place by disregarding the policies applied in the control plane for the selection of the best next-hop for p (the covering prefix). When overlapping prefixes are filtered and packets are forwarded according to the covering prefix, the discrepancy in the routing policies applied both covering and overlapping prefixes can lead to a violation of policies of Internet Service Providing (ISPs) still holding a path towards the overlapping prefix.

This document presents examples of such potential threats, and discusses solutions to the problem. The objective of this draft is to enable the use of prefix filtering while making the routing community aware of the cases where the effects of filtering might turn to be negative for the business of ISPs.

The rest of the document is organized as follows: Section 2 describes some cases in which it is favorable for an AS to filter overlapping prefixes. In Section 3, we provide some scenarios in which the filtering of overlapping prefixes lead to policy violations of other ASes. Section 4 and Section 5 introduce some techniques that ASes can use for, respectively, detect and react to policy violations.

2. Filtering overlapping prefixes

There are different scenarios where filtering an overlapping prefix is relevant to the operations of an AS. In this section, we illustrate examples of these scenarios. We differentiate cases in which the filtering is performed locally from those where the filtering is triggered remotely, by using BGP communities. These scenarios will be used as a base in Section 3 for describing side effects bound with such practices, notably policy violations in the

ASes surrounding the AS applying the procedure.

2.1. Local filtering

Let us first analyze the scenario depicted in Figure 1. AS1 and AS2 are two large autonomous systems spanning a large geographical area and peering in 3 different physical locations. Let AS1 announce prefix 10.0.0.0/22 through the sessions established between the two ASes over all peering links. Additionally, let us define that there is part of AS1's network which exclusively uses prefix 10.0.0.0/24 and which is closer to one specific peering point than to others (right peering link). With the purpose of receiving the traffic from AS2 to prefix 10.0.0.0/24 on the right peering link, AS1 could announce the overlapping prefix on this specific peering point. At the time of the establishment of the peering, it can be defined by both ASes that hot potato routing would happen in both directions of traffic. In this scenario, it becomes relevant for AS2 to enforce such practice by detecting the described situations and automatically issue the appropriate filtering. In this case, by implementing these automatic procedures, AS2 would detect and filter prefix 10.0.0.0/24.

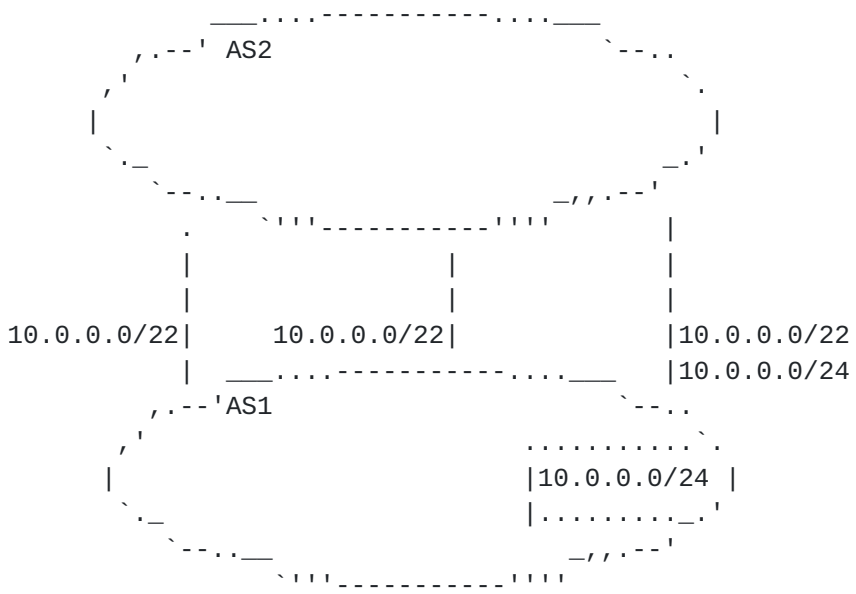

Figure 1: Basic scenario local filtering 1

There are other cases in which there could exist a need for local filtering. For example, a dual homed AS receiving an overlapping prefix from only one of its providers. Figure 2 depicts a simple example of this case.

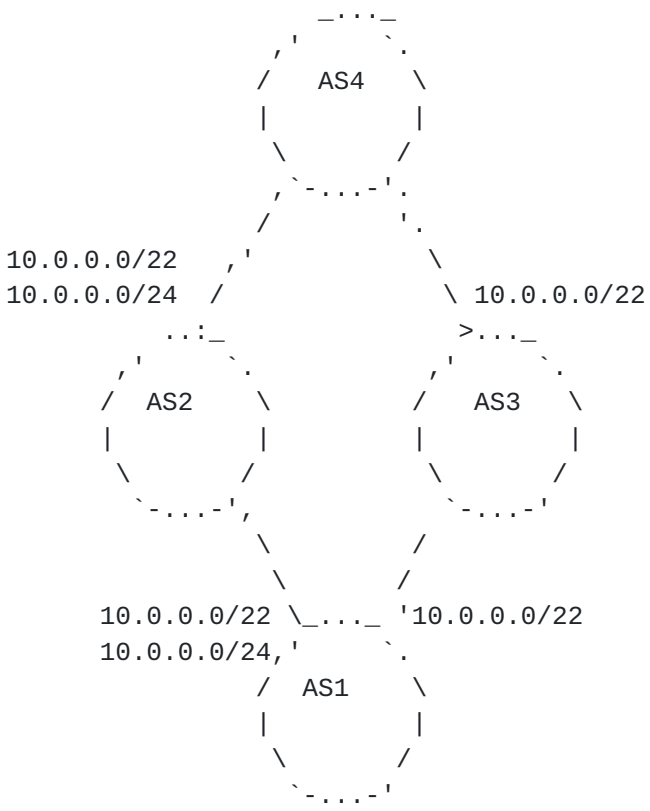

Figure 2: Basic scenario local filtering 2

In this scenario, prefix 10.0.0.0/22 is advertised by AS1 to AS2 and AS3. Both AS propagate the prefix to AS4. Additionally, AS1 advertises prefix 10.0.0.0/24 to AS3, which subsequently propagates the prefix to AS4. 10.0.0.0/22 is a covering prefix for 10.0.0.0/24.

It is possible that AS4 resolves to filter the more specific prefix 10.0.0.0/24. One potential motivation could be the economical preference of the path via AS2 over AS3. Another feasible reason is the existence of a technical policy by AS4 of aggregating incoming prefixes longer than /23.

The above examples illustrate two of the many motivations to configure routing within an AS with the aim of ignoring more specific routes. Operators have reported applying these filters in a manual fashion INIT7-RIPE63 [2]. The relevance of such practice led to investigate automated filtering procedures (DRAFT-WHITE [1]).

2.2. Remotely triggered filtering

ISPs can tag the BGP paths that they propagate to neighboring ASes with communities, so as to tweak the propagation behavior of the ASes

that handle such propagated paths [on_BGP_communities].

Some ISPs allow their direct and indirect customers to use such communities in order to let the receiving AS not export the path to some selected neighboring AS. By combining communities, the prefix could be advertised only to a given peer of the AS providing this feature. Figure 3 illustrates an example of this case.

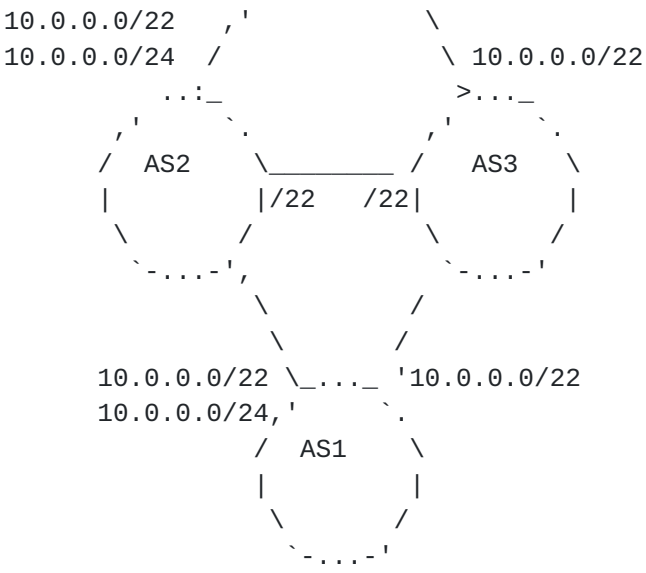

Figure 3: Remote triggered filtering

AS2 and AS3 are peers. Both ASes are providers of AS1. For traffic engineering purposes, AS1 could use communities to prevent AS2 from announcing prefix 10.0.0.0/24 to AS3.

Such technique is useful for operators to tweak routing decisions in order to align with complex transit policies. We will see in the later sections that by producing the same effect as filtering, they can also lead to policy violations at other, distant, ASes.

3. Uses of more specific prefix filtering that violate policies

We describe in this section three configuration scenarios which lead to the violation of the policies of an AS. Note that these examples do not capture all the cases where such policy violation can take place. More examples will be provided in the future revisions of this document.

3.1. Violation caused by Local filtering

In this section we describe cases in which an AS locally filters an overlapping prefix. We show how, depending on the situations of BGP policies, this decision leads to the violation of the policies of neighboring ASes.

3.1.1. Initial setup

We start by describing the basic scenario of this case in Figure 4.

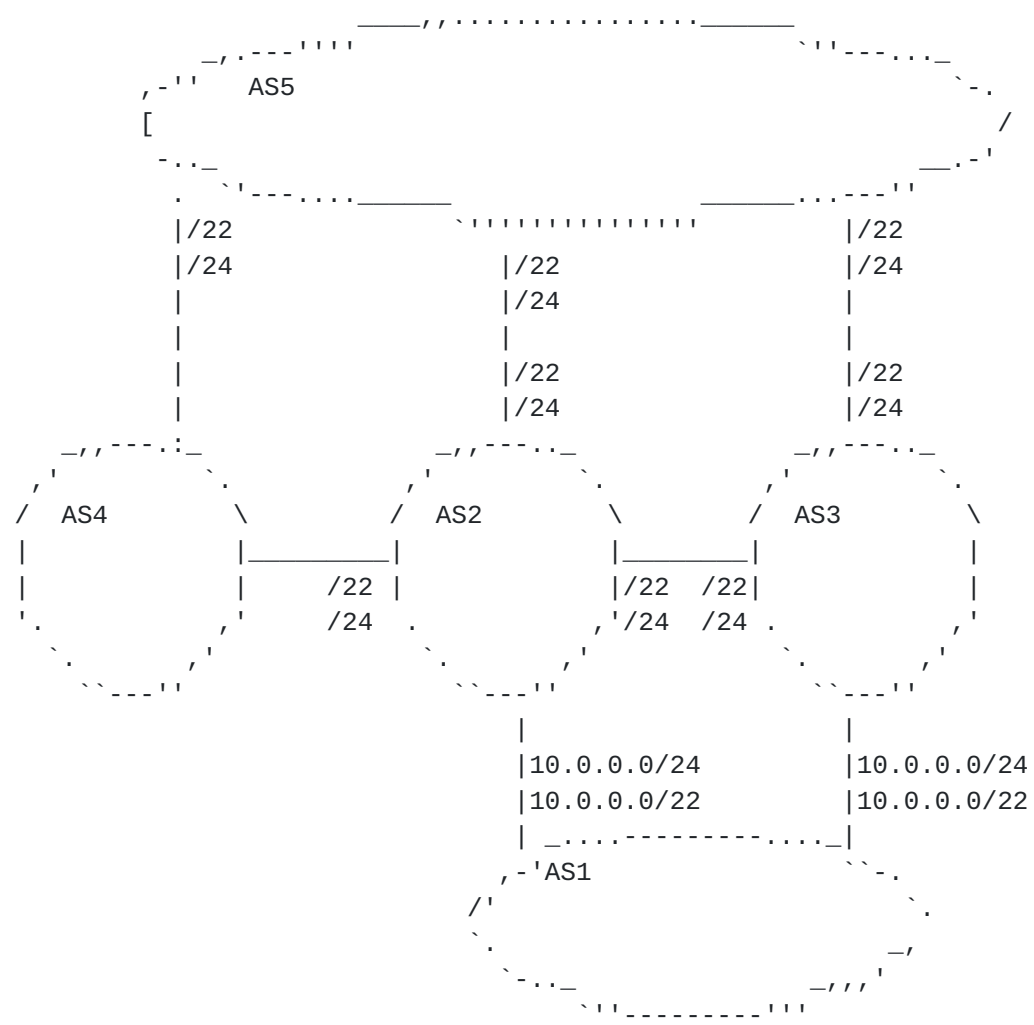

Figure 4: Initial Setup Local

AS1 is a customer of AS2 and AS3. AS2, AS3 and AS4 are customers of AS5. AS2 is establishing a free peering with AS3 and AS4. AS1 is

announcing a covering prefix, 10.0.0.0/22, and an overlapping prefix 10.0.0.0/24 to its providers. In the initial setup, AS2 and AS3 will announce the two prefixes to their peers and transit providers. AS4 receives both prefixes from its peer (AS2) and transit provider (AS5).

3.1.2. Violation of Policy - Case 1

In the next scenarios, we show that if AS4 filters the incoming overlapping prefix from AS5, there is a situation in which the policies of other ASes are violated.

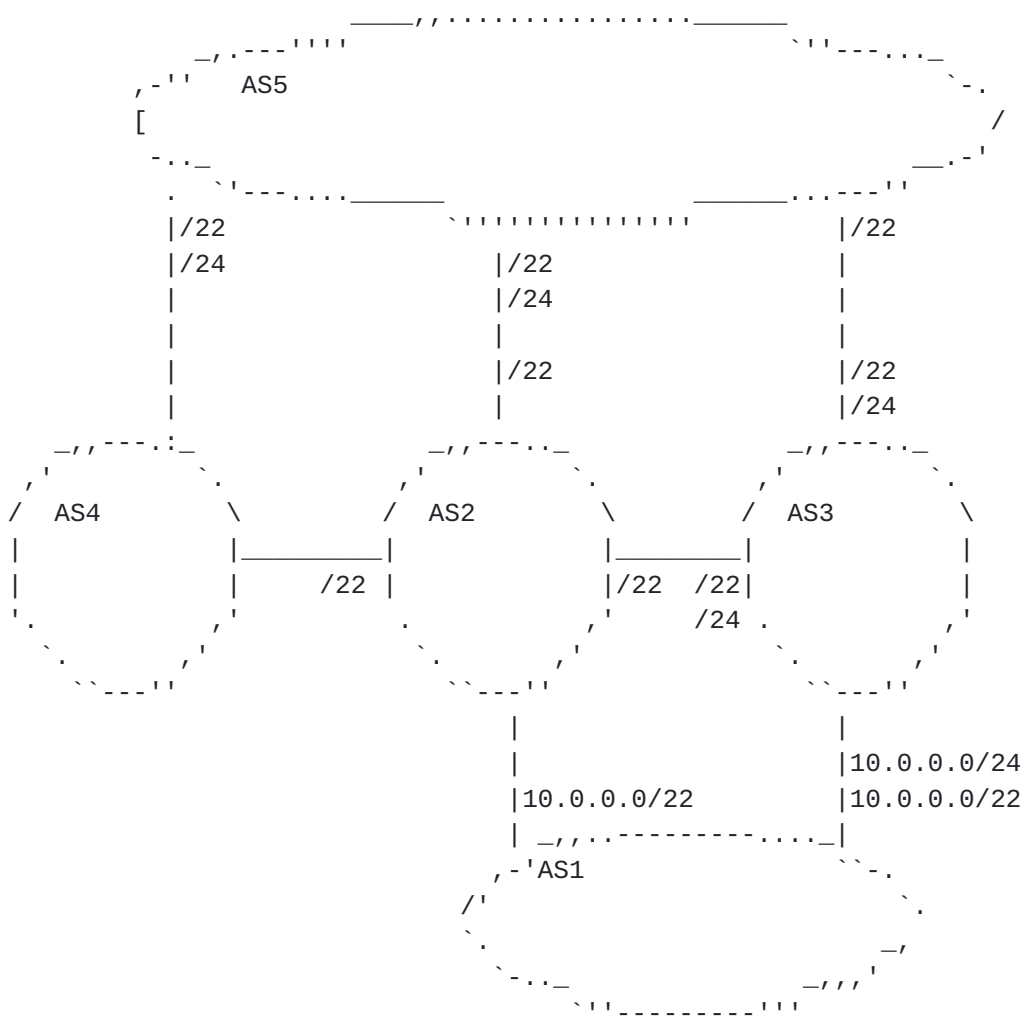

Figure 5: Initial Setup Local

Let us assume the scenario illustrated in Figure 5. For this case, AS1 only propagates the overlapping prefix to AS3. AS4 receives the overlapping prefix only from its traffic provider, AS5.

The described example places AS4 in a situation in which it would be favorable for it to filter the announcement of prefix 10.0.0.0/24 from AS5. Subsequently, traffic originating from AS4 to prefix 10.0.0.0/24 is forwarded to AS2. As AS2 receives the more specific prefix from AS3, traffic originating from AS4 and heading to prefix 10.0.0.0/24 follows the path AS4-AS2-AS3-AS1. This violates the policy of AS2, since it forwards traffic from a peer to a non-customer neighbor.

3.1.3. Violation of Policy - Case 2

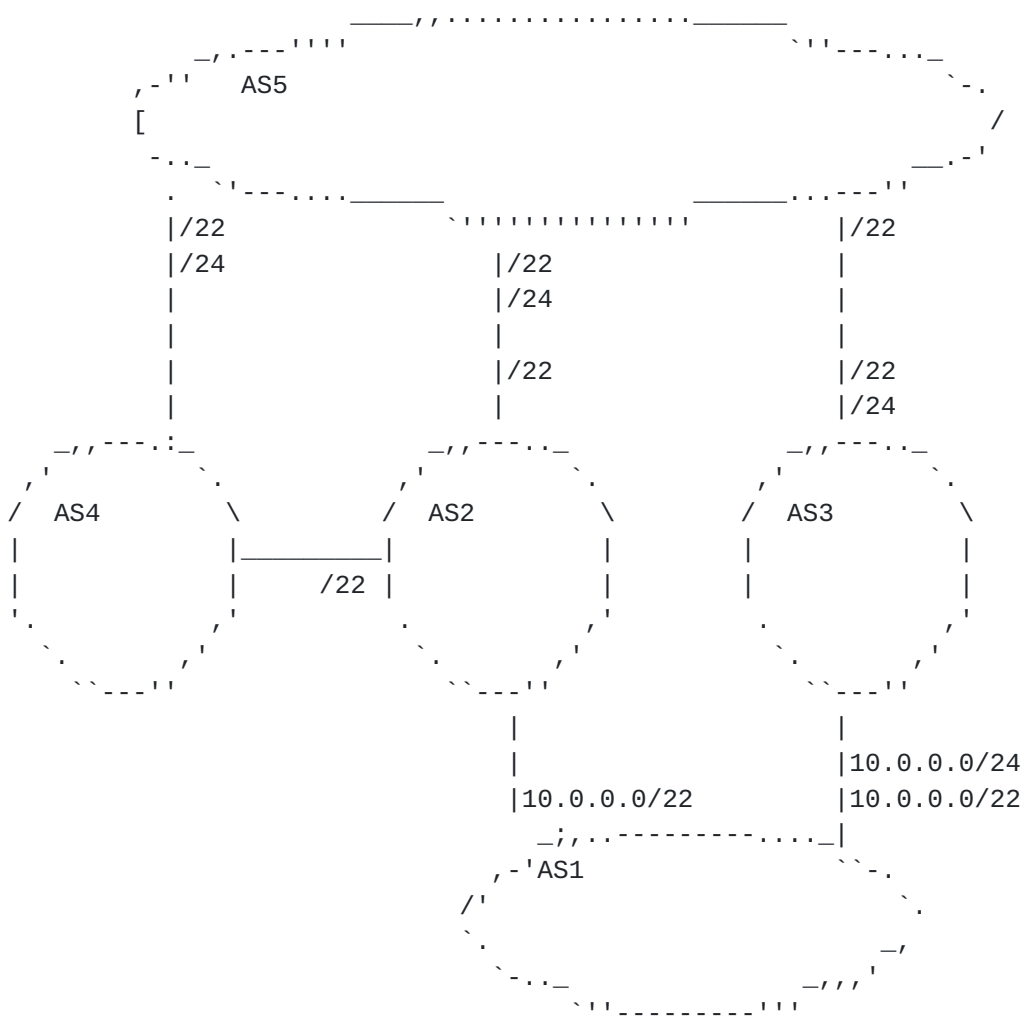

Figure 6: Initial Setup Local

Let us assume a second case where AS2 and AS3 are not peering and AS1

only propagates the overlapping prefix to AS3. AS4 receives the overlapping prefix only from its traffic provider, AS5. This case is illustrated in Figure 6.

Similar to the scenario described in Section 3.1.2, AS4 is in a situation in which it would be favorable to filter the announcement of prefix 10.0.0.0/24 from AS5. Subsequently, traffic originating from AS4 to prefix 10.0.0.0/24 is forwarded to AS2. Traffic originating in AS4 and heading for prefix 10.0.0.0/24 would follow the path AS4-AS2-AS5-AS3-AS1. This path violates the policy of AS2, as this AS is forwarding traffic from a peer to a transit network.

3.2. Violation caused by remotely triggered filtering

We present a configuration scenario in which an AS, using the mechanism described in Section 2.2, informs its provider to selectively announce a covering prefix, leading to the violation of a policy of another AS.

3.2.1. Initial setup

Let AS_cust be a customer AS of AS A and AS B. It owns 10.0.0.0/22, which it advertises through AS A and AS B. Additionally, AS A and AS B are peers.

Both AS A and AS B select their customer path as best, and propagate that path to their customers, providers, and peers.

Some remote ASes will route traffic destined to 10.0.0.0 through (... A Cust 10.0.0.0/24) while some others will route traffic along (... B Cust 10.0.0.0/24).

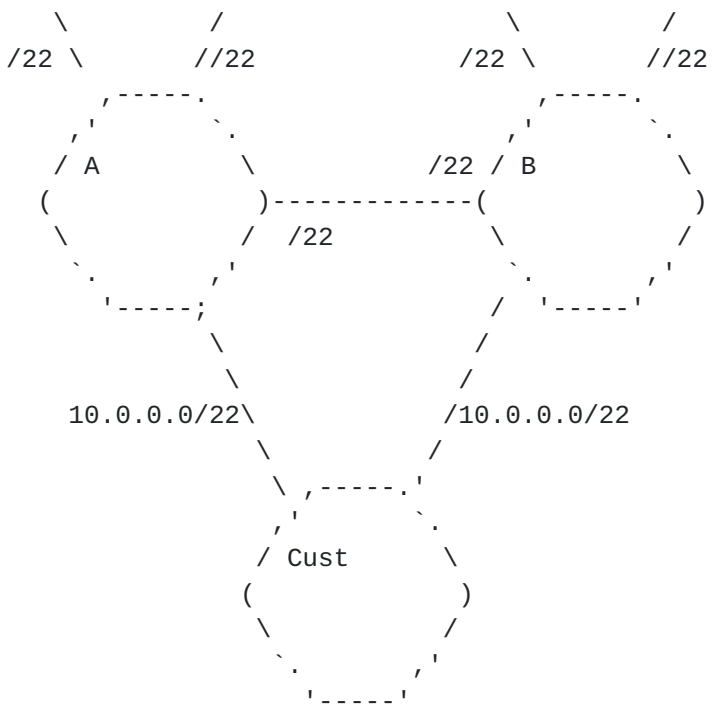

Figure 7: Example scenario

3.2.2. Injection of a more specific

Let AS_cust advertise 10.0.0.0/24 over AS B only. AS B propagates this prefix to its customers, provider and peers, including AS A.

From AS A's point of view, such a path is a "peer path", so that this path will only be advertised to its customers.

All ASes that are not in the customer branch of AS A will receive a path to the /24 that contains AS B, and not AS A, as AS A has not propagated the prefix to other ASes than its customers.

The ASes that are in the customer branch of AS A will receive a path to the /24 that contains AS B and AS A, as AS A has propagated that path to its customers. Some multi-homed customers of ISP A may also receive a path through ISP B, but not through ISP A, from other peering or provider links.

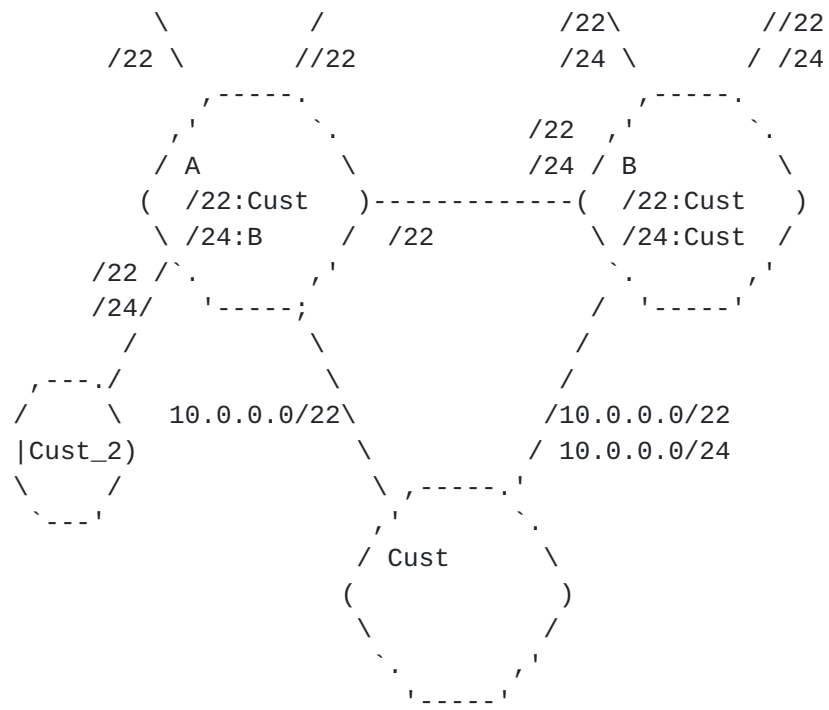

Figure 8: More Specific Injection

Any remote AS that is not lying in the customer branch of A, will receive a path for 10.0.0.0/24 through AS B and not through AS A.

Routing is consistent with usual Internet Routing Policies here, as AS A may only receive traffic destined to 10.0.0.0/24 from its customers, which it forwards to its peer AS B. AS B may receive traffic destined to 10.0.0.0/24 from its customers, providers, and peers, which it directly forwards to its customer AS Cust.

3.2.3. Limiting the scope of the more specific

Now, let us assume that 10.0.0.0/24, which is propagated by AS_Cust to AS B, is tagged so as to have AS B only propagate that path to AS A, using the techniques described in Section 2.2.

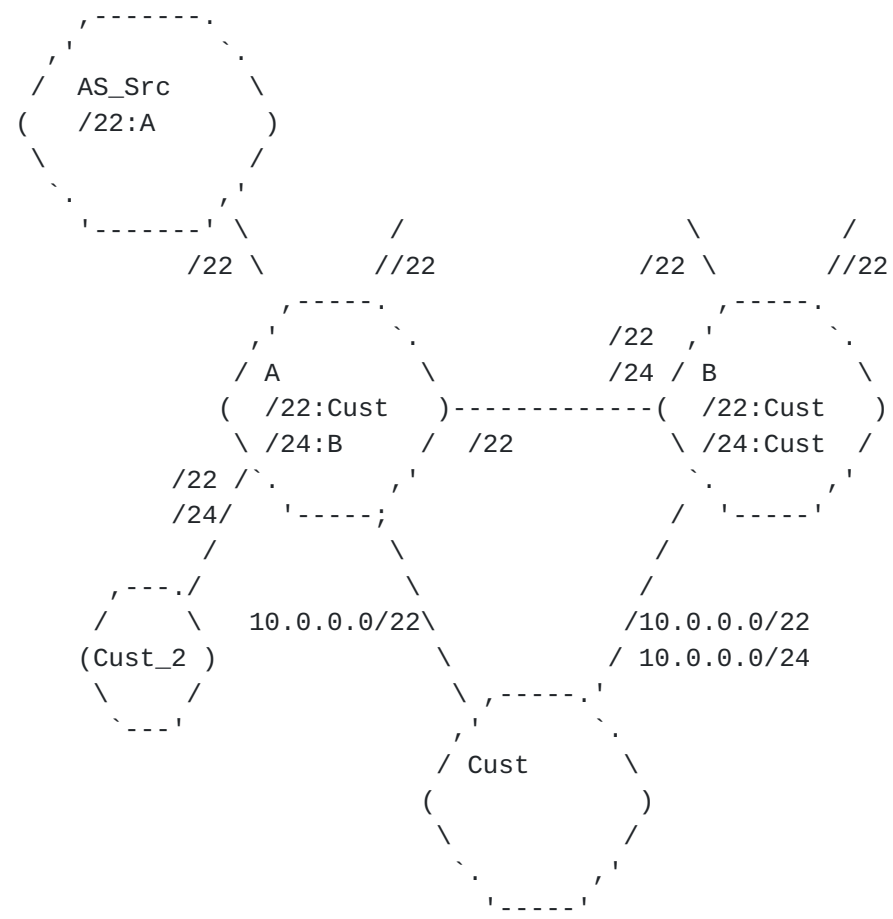

Figure 9: More Specific Injection

From AS A's point of view, such a path is a "peer path", so that this path will only be advertised by AS A to its customers.

All the ASes that are not in the customer branch of AS A nor in the customer branch of AS B will NOT receive a path to 10.0.0.0/24.

All these ASes will forward packets destined to 10.0.0.0/24 according to their routing state for 10.0.0.0/22.

Let us assume that AS_Src is such an AS, and that its best path towards 10.0.0.0/22 is through AS A. In that case, packets sent towards 10.0.0.1 by AS_Src will eventually reach AS A. However, in the dataplane of the nodes of AS A, the longest prefix match for 10.0.0.0 is 10.0.0.0/24, which is reached through AS B, a peer of AS A.

As AS_Src is by definition not in the customer branch of AS A, we are in a situation such that AS A is forwarding non customer originated

traffic along peering links, which violates its policies.

If the path towards 10.0.0.0/24 is propagated by B to its customers, the traffic originated by ASes in the customer branch of AS A will not follow policy-violating data-plane paths as the forwarding of traffic towards these destinations will always be based on FIB entries for 10.0.0.0/24. However, policy-violation can still take place for the traffic originated from all ASes that are neither in the customer branch of A nor in the customer branch of B.

4. Techniques to detect dataplane-based policy violations

We differentiate the techniques available for detecting policy violations from the cases in which the interested AS is the victim or contributor of such operations.

4.1. Being the victim of the policy violation

To detect that its policies have been violated, one ISP can monitor its NetFlow data so as to see if flows entering the ISP network through a non-customer link is being forwarded to a non-customer nexthop.

Detecting such a violation can be done by looking at BGP data to see whether there exists in the RIB a prefix P/p' more specific than P/p such that the nexthop for P/p' is through a peer (or a provider) while P/p is routed through a customer. For each such couple of prefixes, direct communication or looking glasses can be used in order to check whether non-customer neighboring ASes are propagating a path towards P/p (and not towards P/p') to their own customers, peers, or providers. This should trigger a warning as this would mean that ASes in the surrounding area of the current AS are forwarding packets based on the routing entry for the less specific prefix only.

4.2. Being a contributor to the policy violation

It can be considered as problematic to be a contributor of the policy violation as it appears as an abuse of other's network resources.

There may be justifiable reasons for one ISP to perform filtering, either to enforce establishing policies or to provide prefix advertisement scoping features to its customers. These can vary from trouble-shooting purposes to business relationships implementations. Restricting such features for the sake of avoiding contributing to potential policy violations in a peer's network is a bad option.

Netflow data does not help an ISP to detect that it is acting as a contributor of the policy violation. It is thus advisable to obtain as much information as possible of the Internet environment of the AS and assessing the risks of filtering of overlapping prefixes before implementing them.

Monitoring the manipulation of the communities that implement the scoping of prefixes in one's network is recommended to the ISPs which provide these features. The monitored behavior should then be faced against their terms of use.

5. Techniques to counter policy violations

Network Operators can adopt different approaches with respect to policy violation. We classify these actions according to whether they are anticipant or reactive.

Reactive approaches are those in which the operator tries to detect the situations and solves the policy violation through other means than using the routing system.

Anticipant or preventive approaches are those in which the routing system will not let the policy violation actually take place when the configuration scenario is set up.

5.1. Reactive counter-measures

An operator who detects that its policies have been violated can contact the ASes that are likely to have performed the propagation tweaks so as to have them change their behavior.

An operator can account the amount of traffic that has been subject to policy violation, and charge the peer that received the policy-violating traffic. That is, the operator can claim that it has been a provider of that peer for that part of the traffic that transited between the two ASes.

An operator can decide to filter-out the concerned more specific prefix at the peering session over which it was received. In the example of Figure 9, AS A would filter out 10.0.0.0/24 in its eBGP in-filter associated with the eBGP session with AS B. As a result, the traffic destined to that /24 would be forwarded by AS A along its link with AS_Cust, despite the actions performed by AS_Cust to have this traffic coming in through it link with AS B.

5.2. Anticipant counter-measures

5.2.1. Neighbor-specific forwarding

An operator can technically ensure that the traffic destined to a given prefix will be forwarded from an entry point of its AS, only on the basis of the set of paths that have been advertised over that entry point.

5.2.2. Access lists

An operator can configure its routers so as to have them dynamically install an access-list made of the prefixes towards which the forwarding of traffic from that interface would lead to a policy violation. Note that this technique actually lets packets destined to a valid prefix be dropped while they are sent from a neighboring AS that cannot know about the policy violation and hence had no means to avoid the policy violation.

In the example of Figure 9, AS A would install an access-list denying packets matching 10.0.0.0/24 associated with the interface connecting AS_Src. As a result, the traffic destined to that /24 would be dropped, despite the existence of a non policy-violating route towards 10.0.0.0/22.

5.2.3. Automatic filtering

As described in Section 3, filtering of overlapping prefixes can in some scenarios lead to policy violations. Nevertheless, depending on the autonomous system implementing such practice, this operation can in fact prevent these cases. This can be illustrated using the example described in Section 3.1.3: In Figure 6, if AS2 or AS3 filter prefix 10.0.0.0/24, there would be no policy violation for AS2.

6. Conclusions

In this document we described potential threats to policy violation of autonomous systems caused by the filtering of overlapping prefixes by external networks. We provide examples of scenarios of policy violations caused by these practices and introduce some techniques for their detection and counter. We observe that there are reasonable situations in which ASes could filter overlapping prefixes, however, we encourage that network operators implement this type of filters only after considering such threats.

7. References

[on_BGP_communities]
Donnet, B. and O. Bonaventure, "On BGP Communities", ACM SIGCOMM Computer Communication Review vol. 38, no. 2, pp. 55-59, April 2008.

[1] <http://tools.ietf.org/html/draft-white-grow-overlapping-routes-00>

[2] <http://ripe63.ripe.net/presentations/48-How-more-specifics-increase-your-transit-bill-v0.2.pdf>

Authors' Addresses

Juan Camilo Cardona
IMDEA Networks
Avenida del Mar Mediterraneo
Leganes 28919
Spain

Email: juancamilo.cardona@imdea.org

Pierre Francois
IMDEA Networks
Avenida del Mar Mediterraneo
Leganes 28919
Spain

Email: pierre.francois@imdea.org