Network Working Group Camilo Cardona
Internet-Draft Pierre Francois
Intended status: Informational IMDEA Networks
Expires: August 25, 2013 February 21, 2013

Making BGP filtering a habit: Impact on policies
draft-cardona-filtering-threats-01

Abstract

This draft describes 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

Copyright Notice

Table of Contents

1. Introduction

It is common practice for network operators to propagate overlapping prefixes along with the prefixes that they originate. It is also possible for some Autonomous Systems (ASes) to apply different policies to the overlapping (more specific) and the covering (less specific) prefix. Some ASes could even benefit from filtering the overlapping prefixes.

BGP makes independent, policy driven decisions for the selection of the best path to be used for a given IP prefix. However, routers must forward packets using the longest-prefix-match rule, which "precedes" any BGP policy (RFC1812 [4]). Indeed, the existence of a prefix p' that is more specific than a prefix p in the Forwarding Information Base (FIB) will 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 to 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 threats and discusses solutions to the problem. The objective of this draft is to shed light on the use of prefix filtering by 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 discuss some techniques that ASes can use for, respectively, detect and react to policy violations.

2. Filtering overlapping prefixes

There are several scenarios where filtering an overlapping prefix is relevant to the operations of an AS. In this section, we provide examples of these scenarios. We differentiate cases in which the filtering is performed locally from those where the filtering is triggered remotely. 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 a peering point than to others.

To receive the traffic from AS2 to prefix 10.0.0.0/24 on the closer link, AS1 could announce the overlapping prefix only over this specific session. 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 other words, it was agreed that each AS will deliver the traffic to the other AS on the nearest peering link. In this scenario, it becomes relevant to AS2 to enforce such practice by detecting the described situations and automatically issuing the appropriate filtering. In this case, by implementing these automatic procedures, AS2 would detect and filter prefix 10.0.0.0/24.

```
                        ___....-----------....___
                   ,.--' AS2                     `--..
                 ,'                                   `.
                |                                       |
                 `._                                 _.'
                    `--..__                   _,,.--'
                      .    `'''-----------''''       |
                      |                |             |
                      |                |             |
           10.0.0.0/22|     10.0.0.0/22|             |10.0.0.0/22
                      |  ___....-----------....___   |10.0.0.0/24
                    ,.--'AS1                      `--..
                  ,'                        ...........`.
                 |                          |10.0.0.0/24 |
                  `._                       |........._.'
                     `--..__                   _,,.--'
                            `'''-----------''''
```

Figure 1: Basic scenario of 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.

```
                                     _..._
                                  ,'     `.
                                 /   AS4   \
                                 |         |
                                  \       /
                                  ,`-...-'.
                                 /        '.
                 10.0.0.0/22   ,'           \
                 10.0.0.0/24  /              \ 10.0.0.0/22
                           ..:_               >..._
                        ,'     `.           ,'     `.
                       /  AS2    \         /   AS3   \
                       |         |         |         |
                        \       /           \       /
                         `-...-',            `-...-'
                                 \         /
                                  \       /
                       10.0.0.0/22 \_..._ '10.0.0.0/22
                       10.0.0.0/24,'     `.
                                 /  AS1    \
                                 |         |
                                  \       /
                                   `-...-'
```

Figure 2: Basic scenario of local filtering - 2

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

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 [3]. The relevance of such practice led to investigate automated filtering procedures in I-D.WHITE [5].

2.2. Remotely triggered filtering

ISPs can tag the BGP paths that they propagate to neighboring ASes with communities, in order to tweak the propagation behavior of the ASes that handle these paths [1].

Some ISPs allow their direct and indirect customers to use such communities 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.

```
                  10.0.0.0/22   ,'           \
                  10.0.0.0/24  /              \ 10.0.0.0/22
                            ..:_               >..._
                         ,'      `.           ,'     `.
                        /  AS2    \________ /   AS3   \
                        |         |/22   /22|         |
                         \       /           \       /
                          `-...-',            `-...-'
                                  \         /
                                   \       /
                        10.0.0.0/22 \_..._ '10.0.0.0/22
                        10.0.0.0/24,'     `.
                                  /  AS1    \
                                  |         |
                                   \       /
                                    `-...-'
```

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 later sections that by producing the same effect as filtering, they can also lead to policy violations at other, distant, ASes.

3. Uses of overlapping prefix filtering that violate policies

In this section we describe three configuration scenarios that 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 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 that, depending on the BGP policies applied by surrounding ASes, this decision can lead to a policy violation.

3.1.1. Initial setup

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

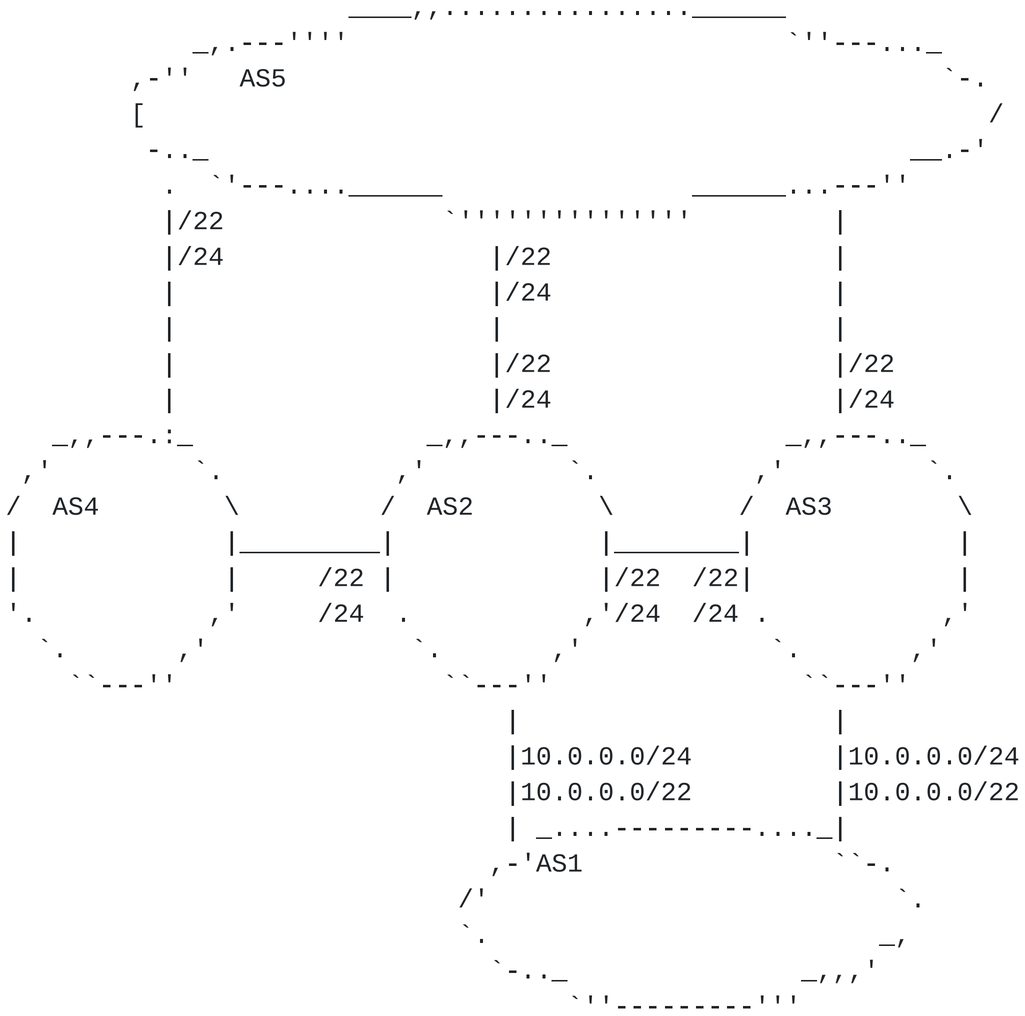

Figure 4: Initial Setup Local

AS1 is a customer of AS2 and AS3. AS2, AS3 and AS4 are customers of AS5. AS2 is establishing a 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). We will consider that AS5 chooses the path through AS3 to reach AS1.

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.

```
                            ____,,................______
                  _,.---''''                            `''---..._
              ,-''   AS5                                          `-.
              [                                                      /
               -.._                                             __.-'
                .  `'---....______                ______...---''
                |/22              `'''''''''''''''         |
                |/24                 |/22                  |
                |                    |/24                  |
                |                    |                     |
                |                    |/22                  |/22
                |                    |                     |/24
         _,,---.:_               _,,---.._              _,,---.._
       ,'         `.           ,'         `.          ,'         `.
      /  AS4        \         /  AS2        \        /  AS3        \
      |             |_________|             |________|             |
      |             |     /22 |             |/22  /22|             |
      '.           ,'          .           ,'     /24 .           ,'
        `.       ,'             `.       ,'            `.       ,'
          ``---''                 ``---''                ``---''
                                      |                    |
                                      |                    |10.0.0.0/24
                                      |10.0.0.0/22         |10.0.0.0/22
                                      | _,,..---------...._|
                                     ,-'AS1                ``-.
                                   /'                          `.
                                   `.                         _,
                                     `-.._               _,,,'
                                          `''---------'''
```

Figure 5: Policy violation after local filtering - Case 1

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 transit 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 from AS4 and heading to prefix 10.0.0.0/24 is forwarded towards AS2. Because AS2 receives the more specific prefix from AS3, traffic 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

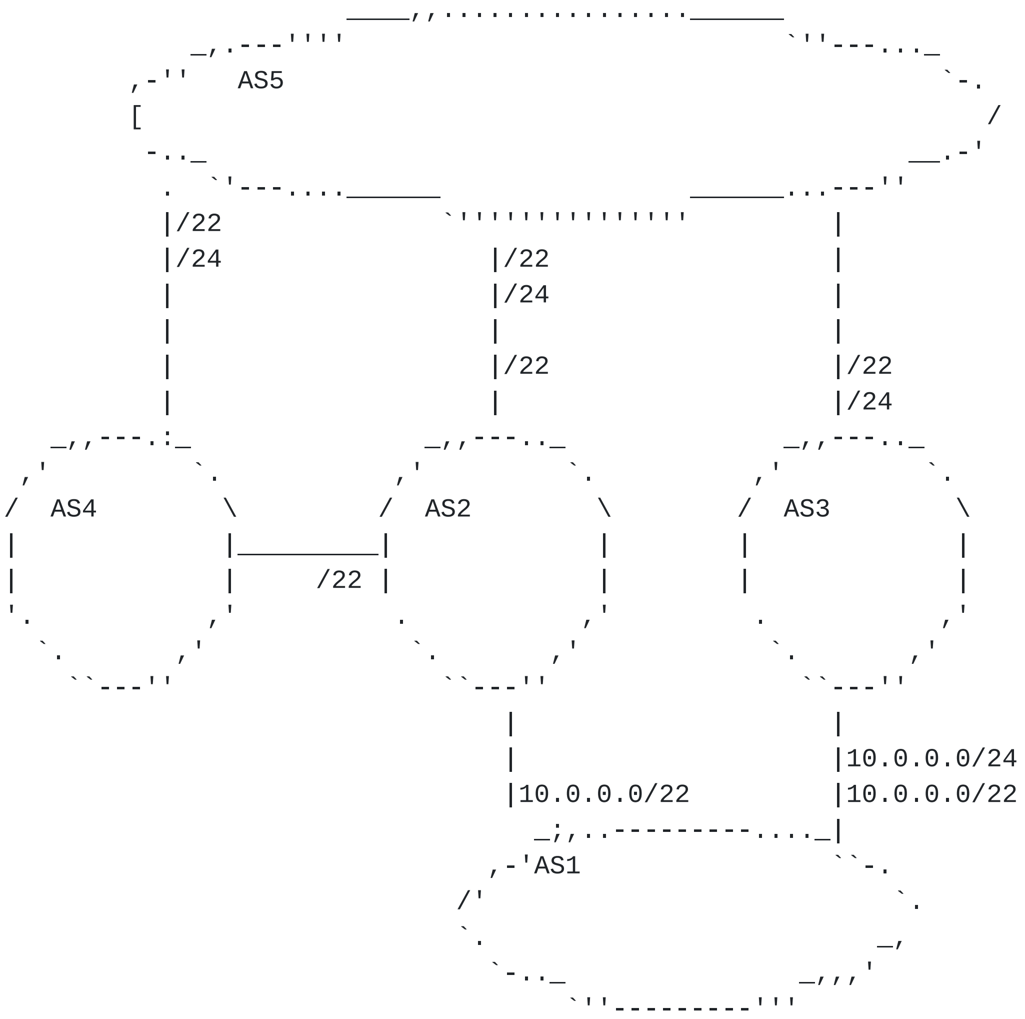

Figure 6: Policy violation after local filtering - Case 2

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 transit 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 from AS4 to prefix 10.0.0.0/24 is forwarded towards AS2. Traffic from AS4 and heading to prefix 10.0.0.0/24 follows 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 an overlapping prefix, leading to the violation of the policy of another AS.

3.2.1. Initial setup

Let AS1 be a customer of AS2 and AS3. AS1 owns 10.0.0.0/22, which it advertises through AS2 and AS3. Additionally, AS2 and AS3 are peers.

Both AS2 and AS3 select their customer path as best, and propagate that path to their customers, providers, and peers. Some remote ASes will route traffic destined to 10.0.0.1 through AS2 while others will route traffic through AS3.

```
                   \         /                  \         /
                /22 \       / /22            /22 \       / /22
                      ,-----.                     ,-----.
                    ,'       `.                 ,'       `.
                   / AS2       \           /22 / AS3       \
                  (             )-------------(             )
                   \           /  /22          \           /
                    `.       ,'                 `.       ,'
                      '-----;                  /  '-----'
                             \                /
                              \              /
                    10.0.0.0/22\            /10.0.0.0/22
                                \          /
                                 \ ,-----.'
                                 ,'       `.
                                / AS1       \
                               (             )
                                \           /
                                 `.       ,'
                                   '-----'
```

Figure 7: Example scenario

3.2.2. Injection of an overlapping prefix

Let AS1 advertise 10.0.0.0/24 over AS3 only. AS3 would propagate this prefix to its customers, providers and peers, including AS2.

From AS2's point of view the path towards 10.0.0.0/24 is a "peer path" and AS2 will only advertise it to its customers. ASes in the customer branch of AS2 will receive a path to the /24 that contains AS3 and AS2. Some multi-homed customers of AS2 may also receive a path through AS3, but not through AS2, from other peering or provider links. Any remote AS that is not lying in the customer branch of AS2, will receive a path for 10.0.0.0/24 through AS3 and not through AS2.

```
                      \         /               /22\         / /22
                   /22 \       / /22            /24 \       /  /24
                         ,-----.                     ,-----.
                       ,'       `.            /22  ,'       `.
                      / AS2       \           /24 / AS3       \
                     (  /22:AS1    )-------------(  /22:AS1    )
                      \ /24:AS3   /  /22          \ /24:AS1   /
                  /22 /`.       ,'                 `.       ,'
                  /24/   '-----;                  /  '-----'
                    /           \                /
              ,---./             \              /
             /     \   10.0.0.0/22\            /10.0.0.0/22
            | AS4   )              \          / 10.0.0.0/24
             \     /                \ ,-----.'
              `---'                ,'       `.
                                   / AS1       \
                                  (             )
                                   \           /
                                    `.       ,'
                                      '-----'
```

Figure 8: Injection of overlapping prefix

AS2 only receives traffic destined to 10.0.0.0/24 from its customers, which it forwards to its peer AS3. Routing is consistent with usual Internet Routing Policies in this case. AS3 could receive traffic destined to 10.0.0.0/24 from its customers, providers, and peers, which it directly forwards to its customer AS1.

3.2.3. Violation of policy by limiting the scope of the overlapping prefix

Now, let us assume that 10.0.0.0/24, which is propagated by AS1 to AS3, is tagged to have AS3 only propagate that path to AS2, using the techniques described in Section 2.2.

```
               ,-------.
             ,'         `.
            /  AS5        \
           (   /22:AS2     )
            \             /
             `.         ,'
               '-------' \         /                  \         /
                      /22 \       //22             /22 \       //22
                            ,-----.                     ,-----.
                          ,'       `.            /22  ,'       `.
                         / AS2       \           /24 / AS3       \
                        (  /22:AS1    )-------------(  /22:AS1    )
                         \ /24:AS3   /  /22          \ /24:AS1   /
                     /22 /`.       ,'                 `.       ,'
                     /24/   '-----;                  /  '-----'
                       /           \                /
                 ,---./             \              /
                /     \   10.0.0.0/22\            /10.0.0.0/22
               (  AS4  )              \          / 10.0.0.0/24
                \     /                \ ,-----.'
                 `---'                 ,'       `.
                                      / AS1       \
                                     (             )
                                      \           /
                                       `.       ,'
                                         '-----'
```

Figure 9: More Specific Injection

From AS2's point of view such a path is a "peer path" and will only be advertised by AS2 to its customers.

All ASes that are not customers of AS2 will not receive a path to 10.0.0.0/24. 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 AS5 is such an AS, and that its best path towards 10.0.0.0/22 is through AS2. Then, packets sent towards 10.0.0.1 by AS5 will eventually reach AS2. However, in the data-plane of the nodes of AS2, the longest prefix match for 10.0.0.1 is 10.0.0.0/24, which is reached through AS3, a peer of AS2. Since AS5 is not in the customer branch of AS2, we are in a situation where AS2 is forwarding non customer originated traffic along peering links, which violates its policies.

4. Techniques to detect 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, an ISP can monitor its traffic data and test if any flow entering the ISP network through a non-customer link is forwarded to a non-customer next-hop.

In the control plane, it is possible for ISPs to identify threats using BGP data. An ISP can seek for overlapping prefixes for which the next-hop is through a provider (or peer), while the next-hop for their covering prefix(es) is through a client. Direct communication or looking glasses can be used to check whether non-customer neighboring ASes are propagating a path towards the covering prefix 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 problematic to be a contributor of a policy violation as it appears as an abuse to the network resources of other ISPs.

There may be justifiable reasons for one ISP to perform filtering, either to enforce established 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 is a bad option.

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

Monitoring the manipulation of the communities that implement the scoping of prefixes is recommended to the ISPs that 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, manually, on a case by case basis.

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.

We will describe these two kind of approaches on the following part of this Section. We will use the scenario depicted in Figure 10 to provide examples for the different techniques.

```
                             ____,,................______
                   _,.---''''                            `''---..._
               ,-''   AS5                                          `-.
              [                                                      /
                -.._                                             __.-'
                 .  `'---....______                ______...---''
                 |/22              `''''''''''''''          |
                 |/24                 |/22                   |
                 |                    |/24                   |
                 |                    |                      |
                 |                    |/22                   |/22
                 |                    |                      |/24
          _,,---.:_               _,,---.._              _,,---.._
        ,'         `.           ,'         `.          ,'         `.
       /  AS4        \         /  AS2        \        /  AS3        \
       |             |_________|             |        |             |
       |             |     /22 |             |        |             |
       '.           ,'          .           ,'         .           ,'
         `.       ,'             `.       ,'            `.       ,'
           ``---''                 ``---''                ``---''
                                       |                     |
                                       |                     |10.0.0.0/24
                                       |10.0.0.0/22          |10.0.0.0/22
                                         _;,..----------...._|
                                      ,-'AS1                ``-.
                                    /'                          `.
                                    `.                         _,
                                      `-.._               _,,,'
                                           `''---------'''
```

Figure 10: Anticipant counter-measures - Base example

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 the traffic that transited between the two ASes.

An operator can decide to filter-out the concerned overlapping prefix at the peering session over which it was received. In the example of Figure 10, AS2 would filter out the incoming prefix 10.0.0.0/24 from the eBGP session with AS5. As a result, the traffic destined to that /24 would be forwarded by AS2 along its link with AS1, despite the actions performed by AS1 to have this traffic coming in through its link with AS3.

5.2. Anticipant counter-measures

5.2.1. Access lists

An operator can configure its routers to 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 10, AS2 would install an access-list denying packets matching 10.0.0.0/24 associated with the interface connecting to AS4. As a result, traffic destined to that prefix would be dropped, despite the existence of a non policy-violating route towards 10.0.0.0/22.

5.2.2. 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 prevent these cases. This can be illustrated using the example described in Figure 10: if AS2 or AS3 filter prefix 10.0.0.0/24, there would be no policy violation for AS2.

5.2.3. Neighbor-specific forwarding

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

As an example, let us analyze the scenario of Figure 10 from the point of view of AS2. The edge router connecting to the AS4 forward packets destined to prefix 10.0.0.0/24 towards AS5. Likewise, it will forward packets destined to prefix 10.0.0.0/22 towards AS1. The router, however, only propagates the path to the covering prefix (10.0.0.0/22) to AS4. An operator could implement the necessary techniques to force the edge router to forward packets coming from AS4 based only on the paths propagated to AS4. Thus, the edge router would forward packets destined to 10.0.0.0/24 towards AS1 in which case no policy violation would occur. This functionality could be implemented in different ways. Check [2] for one particular example of it.

6. Conclusions

In this document we described threats to policies 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

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

[2] Vanbever, L., Francois, P., Bonaventure, O., and J. Rexford, "Customized BGP Route Selection Using BGP/MPLS VPNs", Cisco Systems, Routing Symposium http://www.cs.princeton.edu/~jrex/talks/cisconag09.pdf, October 2009.

[3] "INIT7-RIPE63", <http://ripe63.ripe.net/presentations/48-How-more-specifics-increase-your-transit-bill-v0.2.pdf>.

[4] <http://www.ietf.org/rfc/rfc1812.txt>

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

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