Network Working Group
Internet-Draft
Intended status: Informational
Expires: December 11, 2015

Camilo Cardona
IMDEA Networks/UC3M
Pierre Francois
IMDEA Networks
Paolo Lucente
Cisco Systems
June 9, 2015

Impact of BGP filtering on Inter-Domain Routing Policies
draft-ietf-grow-filtering-threats-06

Abstract

This document describes how unexpected traffic flows can emerge across an autonomous system, as the result of other autonomous systems filtering, or restricting the propagation of more specific prefixes. We provide a review of the techniques to detect the occurrence of this issue and defend against it.

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 December 11, 2015.

Copyright Notice

Copyright (c) 2015 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (http://trustee.ietf.org/license-info) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must

include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

1. Introduction 2
1.1. Terminology 3
2. Unexpected Traffic Flows 4
2.1. Local filtering 4
2.1.1. Unexpected traffic flows caused by local filtering of more specific prefixes 5
2.2. Remote filtering 6
2.2.1. Unexpected traffic flows caused by remotely triggered filtering of more specific prefixes 7
3. Techniques to detect unexpected traffic flows caused by filtering of more specific prefixes 8
3.1. Existence of unexpected traffic flows within an AS . . . 8
3.2. Contribution to the existence of unexpected traffic flows in another AS 9
4. Techniques to Traffic Engineer unexpected traffic flows . . . 10
4.1. Reactive Traffic Engineering 11
4.2. Proactive measures 12
4.2.1. Access lists 12
4.2.2. Neighbor-specific forwarding 13
5. Conclusions 14
6. Security Considerations 14
7. IANA Considerations 14
8. Acknowledgments 14
9. References 14
9.1. Informative References 14
9.2. URIs 15
Authors' Addresses 15

1. Introduction

It is common practice for network operators to propagate a more specific prefix in the BGP routing system, along with the less specific prefix that they originate. It is also possible for some Autonomous Systems (ASes) to apply different policies to the more specific and the less specific prefix.

While BGP makes independent, policy driven decisions for the selection of the best path to be used for a given IP prefix, routers must forward packets using the longest-prefix-match rule, which "precedes" any BGP policy (RFC1812 [1]). 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 more specific prefix). This process takes place by disregarding the policies applied in the control plane for the selection of the best next-hop for p'. When an Autonomous System filters more specific prefixes and forwards packets according to the less specific prefix, the discrepancy in the routing policies applied to the less and the more specific prefixes can create unexpected traffic flows that infringe the policies of other ASes, still holding a path towards the more specific prefix.

The objective of this draft is to shed light on possible side effects associated with more specific prefix filtering. This document presents examples of such side effects and discusses approaches towards solutions to the problem.

The rest of the document is organized as follows: In Section 2 we provide some scenarios in which the filtering of more specific prefixes leads to the creation of unexpected traffic flows. Section 3 and Section 4 discuss some techniques that ASes can use for, respectively, detect and react to unexpected traffic flows. We conclude in Section 5.

1.1. Terminology

More specific prefix: A prefix in the routing table with an address range that is covered by a shorter prefix also present in the routing table.

Less specific prefix: A prefix in the routing table with an address range partially covered by other prefixes.

We re-use the definitions of customer-transit peering and settlement-free peering of RFC4384 [2].

Selective advertisement: The behavior of only advertising a self originated BGP path for a prefix over a strict subset of the eBGP sessions of the AS.

Selective propagation: The behavior of only propagating a BGP path for a prefix over a strict subset of the eBGP sessions of an AS.

Local filtering: The behavior of explicitly ignoring a BGP path received over an eBGP session.

Remote filtering: The behavior of triggering selective propagation of a BGP path at a distant AS. Note that this is typically achieved by tagging a self-originated path with BGP communities defined by the distant AS.

Unexpected traffic flow: Traffic flowing between two neighboring ASes of an AS, although the transit policy of that AS is to not provide connectivity between these two neighbors. A traffic flow across an AS, between two of its transit providers, or between a transit provider and one of its settlement-free peers, are classical examples of unexpected traffic flows.

2. Unexpected Traffic Flows

In this section, we describe how more specific prefix filtering can lead to unexpected traffic flows in other, remote, ASes. We differentiate cases in which the filtering is performed locally from those where the filtering is triggered remotely.

2.1. Local filtering

Local filtering can be motivated by different reasons, such as: (1) Traffic engineering, where an AS wants to control its local outbound traffic distribution using only the policy applied to the less specific prefix. Such a practice was notably documented in [3] (2) Enforcing contract compliance, where, for instance, an AS avoids a settlement-free peer to attract traffic to one link by using selective advertisement, when this is not allowed by their peering agreement. (3) The need for Forwarding Information Base memory preservation sometimes pushes ISP operators to filter more specific prefixes.

Figure 1 illustrates a scenario in which one AS is motivated to perform local filtering due to outbound traffic engineering. The figure depicts AS64504, and two of its neighboring ASes, AS64502, and AS64505. AS64504 has a settlement-free peering with AS64502 and is a customer of AS64505. AS64504 receives from AS64505 prefixes 2001:DB8::/32 and 2001:DB8::/34, a less specific and a more specific prefix, respectively. AS64504 receives only the less specific prefix 2001:DB8::/32 from AS64502.



Figure 1: Local Filtering

Due to economic reasons, AS64504 might prefer to send traffic to AS64502 instead of AS64505. However, even if paths received from AS64502 are given a large local preference, routers in AS64504 will still send traffic to prefix 2001:DB8::/34 via neighbor AS64505. This situation may push AS64504 to apply an inbound filter for the more specific prefix, 2001:DB8::/34, on the session with AS64505. After the filter is applied, traffic to the more specific prefix will be sent to AS64502

2.1.1. Unexpected traffic flows caused by local filtering of more specific prefixes

In this section, we show how the decision of AS64504 to perform local filtering creates unexpected traffic flows in AS64502. Figure 2 shows the whole picture of the scenario; where AS64501 is a customer of AS64503 and AS64502. AS64503 is a settlement-free peer with AS64502. AS64503 and AS64502 are customers of AS64505. The AS originating the two prefixes, AS64501, performs selective advertisement with the more specific prefix and only announces it to AS64503.

After AS64504 locally filters the more specific prefix, traffic from AS64504 to prefix 2001:DB8::/34 is forwarded towards AS64502. Because AS64502 receives the more specific prefix from AS64503, traffic from AS64504 to 2001:DB8::/34 follows the path AS64504-AS64502-AS64503-AS64501. AS64502's BGP policies are implemented to avoid transporting traffic between AS64504 and AS64503. However, due to the discrepancies of routes between the more and the less specific prefixes, unexpected traffic flows between AS64504 and AS64503 exist in AS64502's network.

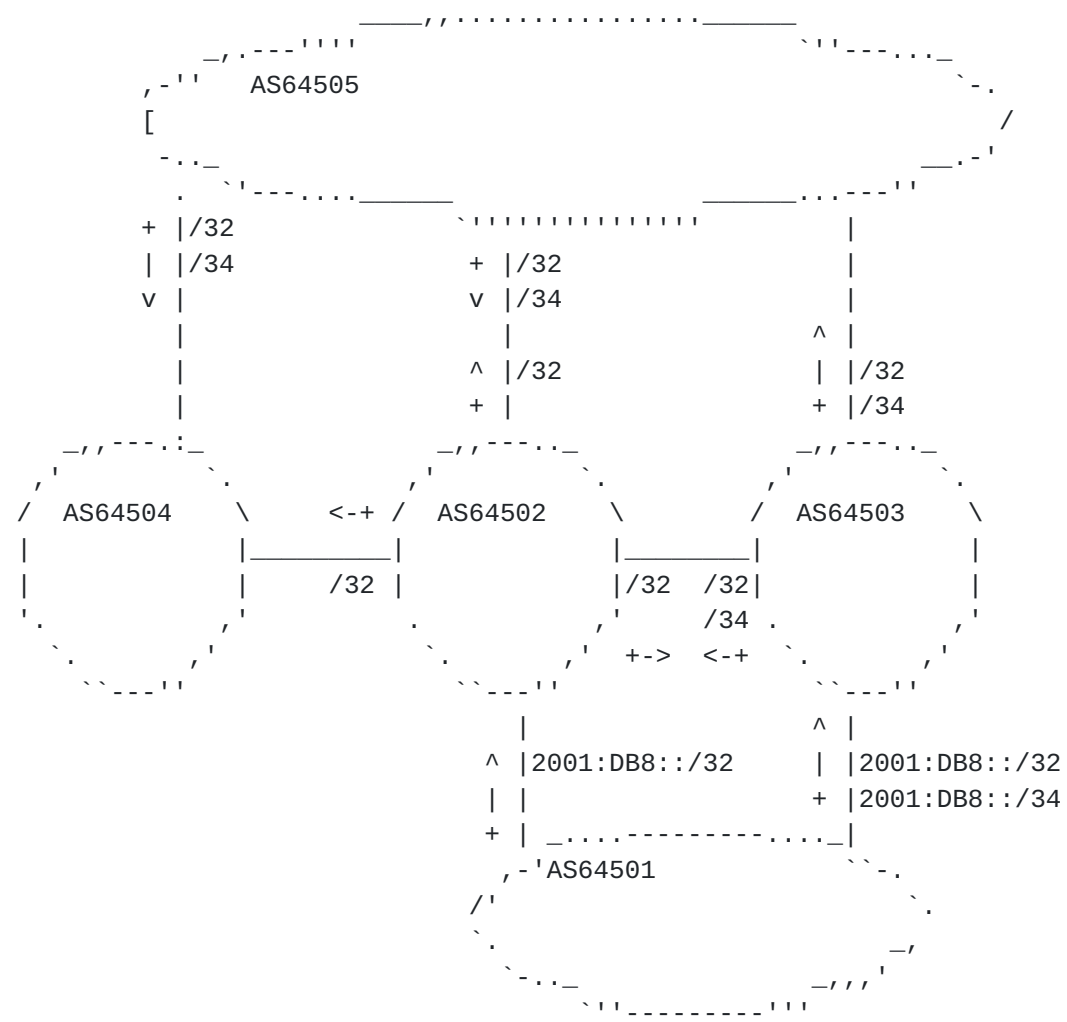


Figure 2: Unexpected traffic flows due to local filtering

2.2. Remote 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 customers to use such communities to let the receiving AS not export the path to some selected neighboring ASes. By combining communities, the prefix could be advertised only to a given peer of the AS providing this feature. Remote filtering can be leveraged by an AS to, for instance, limit the scope of prefixes and hence perform a more granular inbound traffic engineering.

Figure 3 illustrates a scenario in which an AS uses BGP communities to command its provider to selectively propagate a more specific prefix. Let AS64501 be a customer of AS64502 and AS64503. AS64501

originates prefix 2001:DB8::/32, which it advertises to AS64502 and AS64503. AS64502 and AS64503 are settlement-free peers. Let AS64501 do selective advertisement and only propagate 2001:DB8::/34 over AS64503. AS64503 would normally propagate this prefix to its customers, providers, and peers, including AS64502.

Let us consider that AS64501 decides to limit the scope of the more specific prefix. AS64501 can make this decision based on its traffic engineering strategy. To achieve this, AS64501 can tag the more specific prefix with a set of communities that leads AS64503 to only propagate the path to AS64502.

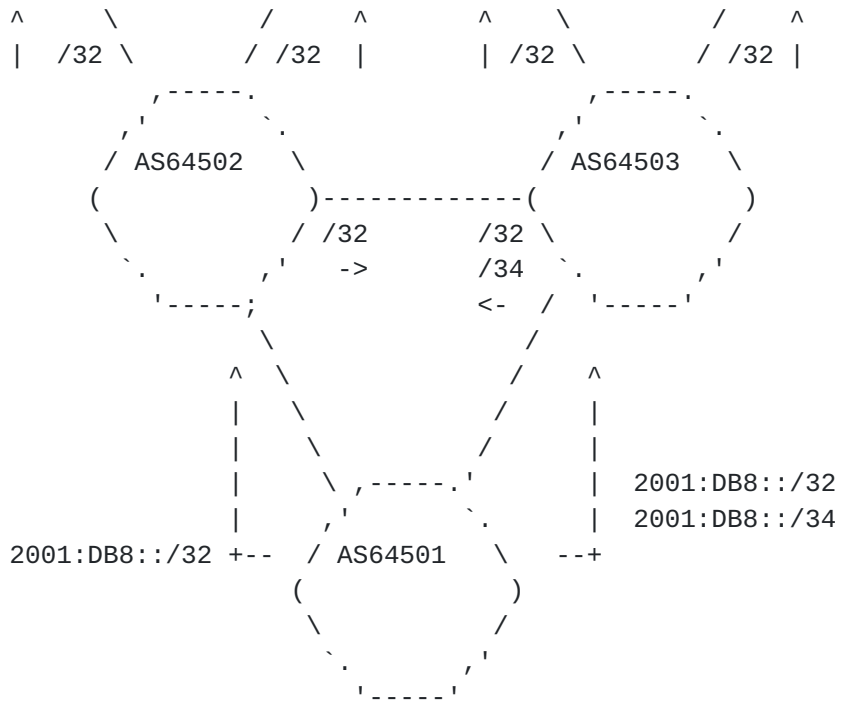


Figure 3: Remote triggered filtering

2.2.1. Unexpected traffic flows caused by remotely triggered filtering of more specific prefixes

Figure 4 expands the scenario from Figure 3 and includes other AS peering with ASes 64502 and 64503. Due to the limitation on the scope performed on the more specific prefix, ASes that are not customers of AS64502 will not receive a path for 2001:DB8::/34. These ASes will forward packets destined to 2001:DB8::/34 according to their routing state for 2001:DB8::/32. Let us assume that AS64505 is such an AS, and that its best path towards 2001:DB8::/32 is through AS64502. Packets sent towards 2001:DB8::1 by AS64505 will reach AS64502. However, in the data-plane of the nodes of AS64502, the longest prefix match for 2001:DB8::1 is 2001:DB8::/34, which is

reached through AS64503, a settlement-free peer of AS64502. Since AS64505 is not in the customer branch of AS64502, we are in a situation in which traffic flows between non-customer ASes take place in AS64502.

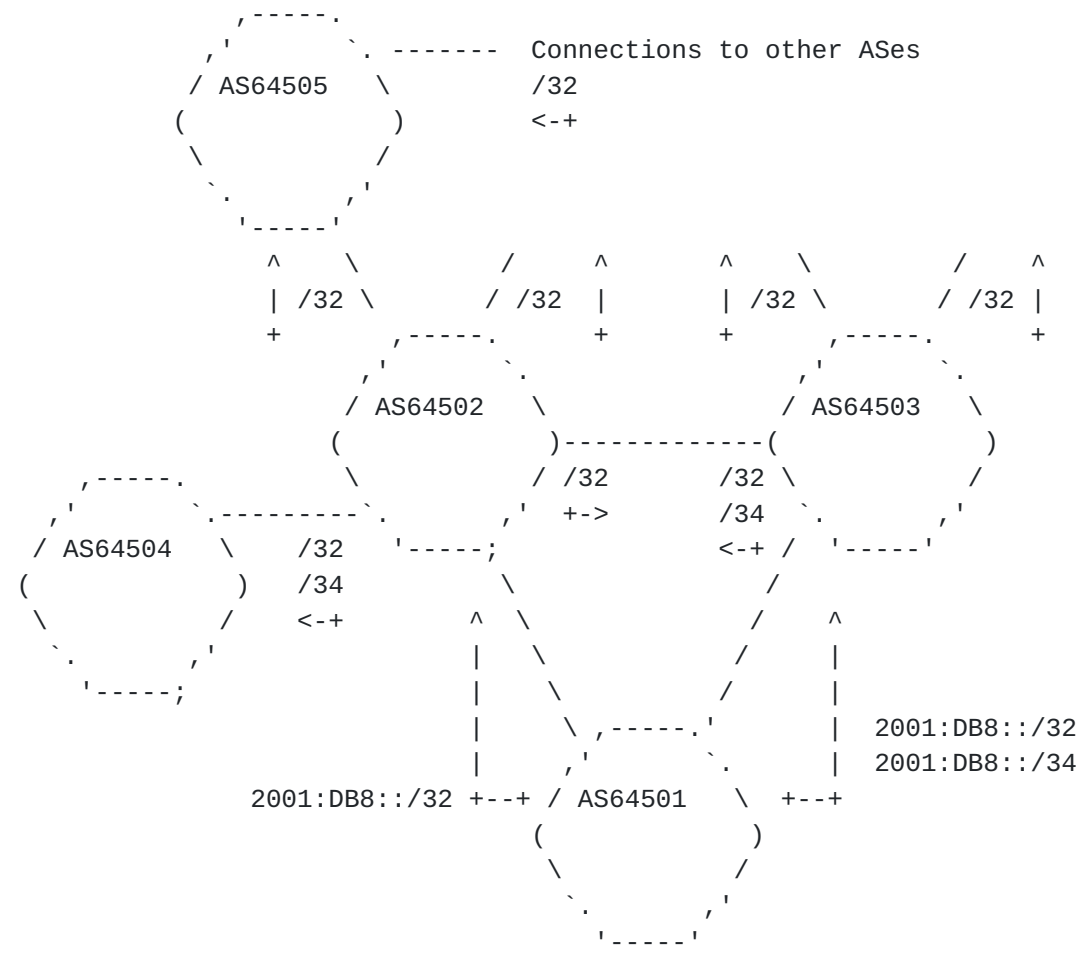


Figure 4: Unexpected traffic flows due to remote triggered filtering

3. Techniques to detect unexpected traffic flows caused by filtering of more specific prefixes

3.1. Existence of unexpected traffic flows within an AS

To detect if unexpected traffic flows are taking place in its network, an ISP can monitor its traffic data to check if it is providing transit between two of its peers, although his policy is configured to not provide such transit. IPFIX (RFC7011 [3]) is an example of a technology that can be used to export information regarding traffic flows across the network. Traffic information must be analyzed under the perspective of the business relationships with

neighboring ASes. Open source tools such as [4] can be used to this end.

Note that the AS detecting the unexpected traffic flow may simply realize that his policy configuration is broken. The first recommended action upon detection of an unexpected traffic flow is to verify the correctness of the BGP configuration.

Once it has been assessed that the local configuration is correct, the operator should check if the unexpected flow arose due to filtering of BGP paths for more-specific prefixes by neighboring ASes. This can be performed in two steps. First, the operator should check whether the neighboring AS originating the unexpected flow is forwarding traffic using a less-specific prefix that is announced to it by the affected network. The second step is to try to infer the reason why the neighboring AS does not use the more-specific path for forwarding, i.e., finding why the more-specific prefix was filtered. We note that due to the distributed nature and restricted visibility of the steering of BGP policies, this second step is deemed to not identify the origin of the problem with guaranteed accuracy.

For the first step, the operator should check if the destination address of the unexpected traffic flow is locally routed as per a more specific prefix only received from non-customer peers. The operator should also check if there are paths to a less specific prefix received from a customer, and hence propagated to peers. If these two situations happen at the same time, the neighboring AS at the entry point of the unexpected flow is routing the traffic based on the less specific prefix, although the ISP is actually forwarding the traffic via non-customer links.

For the second step, human interaction or looking glasses can be used in order to find out whether local filtering, remote filtering, or selective propagation was performed on the more specific prefix. We are not aware, at the time of this writing, of any openly available tool that can automatically perform this operation.

3.2. Contribution to the existence of unexpected traffic flows in another AS

It can be considered problematic to be causing unexpected traffic flows in other ASes. It is thus advisable for an AS to assess the risks of filtering more specific prefixes before implementing them by obtaining as much data information as possible about its surrounding routing environment.

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 relationship implementations. Restricting the use of these features for the sake of avoiding the creation of unexpected traffic flows is not a practical option.

In order to assess the risk of filtering more specific prefixes, the AS would need information on the routing policies and the relationships among external ASes, to detect if its actions could trigger the appearance of unexpected traffic flows. With this information, the operator could detect other ASes receiving the more specific prefix from non-customer ASes, while announcing the less specific prefix to other non-customer ASes. If the filtering of the more specific prefix leads other ASes to send traffic for the more specific prefix to these ASes, an unexpected traffic flow can arise. However, the information required for this operation is difficult to obtain, due to the distributed nature of BGP policies. We are not aware, at the time of this writing, of any openly available tool that can automatically perform this procedure.

4. Techniques to Traffic Engineer unexpected traffic flows

Network Operators can adopt different approaches with respect to unexpected traffic flows. Note that due the complexity of inter-domain routing policies, there is not a single solution that can be applied to all situations. We provide potential solutions that ISPs must evaluate against each particular case. We classify these actions according to whether they are proactive or reactive.

Reactive approaches are those in which the operator tries to detect the situations via monitoring and solve unexpected traffic flows, manually, on a case-by-case basis.

Anticipant or preventive approaches are those in which the routing system will not let the unexpected traffic flows actually take place when the scenario arises.

We use the scenario depicted in Figure 5 to describe these two kinds of approaches. Based on our analysis, we observe that proactive approaches can be complex to implement and can lead to undesired effects. Therefore, we conclude that the reactive approach is the more reasonable recommendation to deal with unexpected flows.

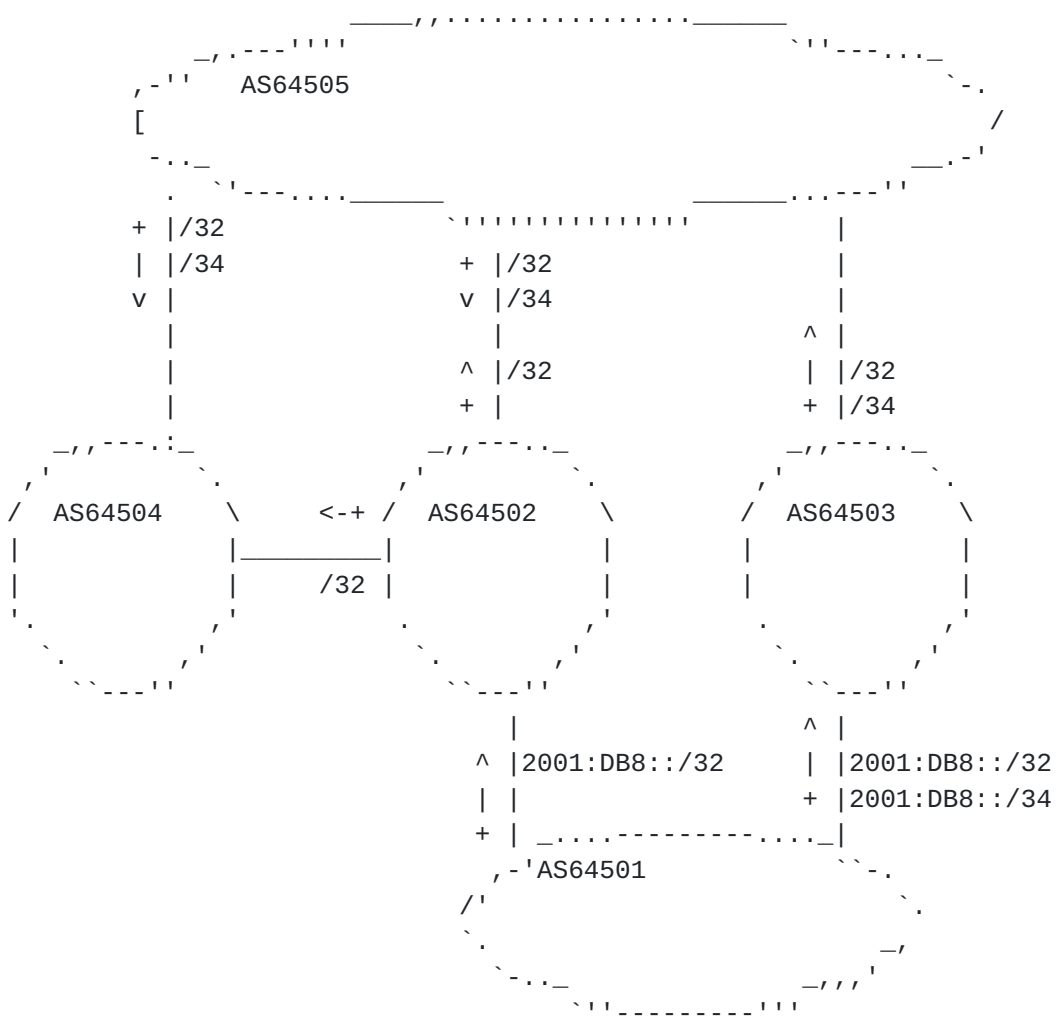


Figure 5: Traffic Engineering of unexpected traffic flows - Base example

4.1. Reactive Traffic Engineering

An operator who detects unexpected traffic flows originated by any of the cases described in Section 2 can contact the ASes that are likely to have performed the propagation tweaks, inform them of the situation, and persuade them to change their behavior.

If the situation remains, the operator can implement prefix filtering in order to stop the unexpected flows. The operator can decide to perform this action over the session with the operator announcing the more specific prefix or over the session with the neighboring AS from which it is receiving the traffic. Each of these options carry a different repercussion for the affected AS. We describe briefly the two alternatives.

- An operator can decide to stop announcing the less specific prefix at the peering session with the neighboring AS from which it is receiving traffic to the more specific prefix. In the example of Figure 5, AS64502 would filter out the prefix 2001:DB8::/32 from the eBGP session with AS64504. In this case, traffic heading to the prefix 2001:DB8::/32 from AS64501 would no longer traverse AS64502. AS64502 should evaluate if solving the issues originated by the unexpected traffic flows are worth the loss of this traffic share.

- An operator can decide to filter out the more specific prefix at the peering session over which it was received. In the example of Figure 5, AS64502 would filter out the incoming prefix 2001:DB8::/34 from the eBGP session with AS64505. As a result, the traffic destined to that /32 would be forwarded by AS64502 along its link with AS64501, despite the actions performed by AS64501 to have this traffic coming in through its link with AS64503. However, as AS64502 will no longer know a route to the more specific prefix, it risks losing the traffic share from customers different from AS64501 to that prefix. Furthermore, this action can generate conflicts between AS64502 and AS64501, since AS64502 does not follow the routing information expressed by AS64501 in its BGP announcements.

It is possible that the behavior of the neighboring AS causing the unexpected traffic flows opposes the peering agreement. In this case, an operator could account the amount of traffic that has been subject to the unexpected flows, using traffic measurement protocols such as IPFIX, and charge the peer for that 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.

4.2. Proactive measures

4.2.1. Access lists

An operator could install access-lists to prevent unexpected traffic flows from happening in the first place. In the example of Figure 5, AS64502 would install an access-list denying packets matching 2001:DB8::/34 associated with the interface connecting to AS64504. As a result, traffic destined to that prefix would be dropped, despite the existence of a valid route towards 2001:DB8::/32.

The operational overhead of such a solution is considered high, as the operator would have to constantly adapt these access-lists to accommodate inter-domain routing changes. Moreover, this technique lets packets destined to a valid prefix be dropped while they are sent from a neighboring AS that may not know about the policy

conflict, and hence had no means to avoid the creation of unexpected traffic flows. For this reason, this technique can be considered harmful and is thus not recommended for implementation.

4.2.2. 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 5 from the point of view of AS64502. The edge router connecting to the AS64504 forwards packets destined to prefix 2001:DB8::/34 towards AS64505. Likewise, it forwards packets destined to prefix 2001:DB8::/32 towards AS64501. The router, however, only propagates the path of the less specific prefix (2001:DB8::/32) to AS64504. An operator could implement the necessary techniques to force the edge router to forward packets coming from AS64504 based only on the paths propagated to AS64504. Thus, the edge router would forward packets destined to 2001:DB8::/34 towards AS64501 in which case no unexpected traffic flow would occur.

Different techniques could provide this functionality; however, their technical implementation can be complex to design and operate. An operator could, for instance, employ Virtual Routing Forwarding (VRF) tables (RFC4364 [4]) to store the routes announced to a neighbor and forward traffic exclusively based on those routes. [2] describes the use of such an architecture for Internet routing, and provides a description of its limitations.

In such architecture, packets received from a peer would be forwarded solely based on the paths that fit the path propagation policy for that peer, and not based on the global routing table of the router. As a result, a more specific path that would not be propagated to a peer will not be used to forward a packet from that peer, and the unexpected flow will not take place. Packets will be forwarded based on the policy compliant less specific prefix. However, note that an operator must make sure that all their routers could support the potential performance impact of this approach.

Note that similarly to the solution described in Section 4.1, this approach could create conflicts between AS64502 and AS64501, since the traffic forwarding performed by AS64502 goes against the policy of AS64501.

5. Conclusions

In this document, we described how the filtering of more specific prefixes can potentially create unexpected traffic flows in remote ASes. We provided examples of scenarios in which unexpected traffic flows are caused by these practices and introduce some techniques for their detection and prevention. Analyzing the different options for dealing with this kind of problems, we recommend ASes to implement monitoring systems that can detect them and react to them according to the specific situation. Although we observe that there are reasonable situations in which ASes could filter more specific prefixes, we encourage network operators to implement this type of filters considering the cases described in this document.

6. Security Considerations

It is possible for an AS to use any of the methods described in this document to deliberately reroute traffic flowing through another AS. The objective of this document is to inform on this potential routing security issue, and analyse ways for operators to defend against them.

It must be noted that, at the time of this document, there are no existing or proposed tools to automatically protect against such behavior. Network monitoring allowing for the detection of unexpected traffic flows exist, but automated configuration changes to solve the problem do not.

7. IANA Considerations

This document has no IANA actions.

8. Acknowledgments

The authors would like to thank Wes George, Jon Mitchell, and Bruno Decraene for their useful suggestions and comments.

9. References

9.1. Informative 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] "pmacct project: IP accounting iconoclasm", <http://www.pmacct.net>.

9.2. URIs

[1] http://www.ietf.org/rfc/rfc1812.txt

[2] http://www.ietf.org/rfc/rfc4384.txt

[3] http://www.ietf.org/rfc/rfc7011.txt

[4] http://www.ietf.org/rfc/rfc4364.txt

Authors' Addresses

Camilo Cardona
IMDEA Networks/UC3M
Avenida del Mar Mediterraneo, 22
Leganes 28919
Spain

Email: juancamilo.cardona@imdea.org

Pierre Francois
IMDEA Networks
Avenida del Mar Mediterraneo, 22
Leganes 28919
Spain

Email: pierre.francois@imdea.org

Paolo Lucente
Cisco Systems
170 W. Tasman Drive
San Jose, CA 95134
USA

Email: plucente@cisco.com