Network Working Group
Internet-Draft
Intended status: Informational
Expires: August 17, 2014

Camilo Cardona
IMDEA Networks/UC3M
Pierre Francois
IMDEA Networks
Paolo Lucente
Cisco Systems

February 13, 2014

Making BGP filtering a habit: Impact on policies
draft-ietf-grow-filtering-threats-02

Abstract

Network operators define their BGP policies based on the business relationships that they maintain with their peers. By limiting the propagation of BGP prefixes, an autonomous system avoids the existence of flows between BGP peers that do not provide any economical gain. This draft describes how unexpected traffic flows can emerge in autonomous systems due to the filtering of overlapping BGP prefixes by 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 August 17, 2014.

Copyright Notice

Copyright (c) 2014 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
2. Filtering overlapping prefixes 3
2.1. Local filtering 3
2.2. Remotely triggered filtering 6
3. Uses of overlapping prefix filtering that create unexpected traffic flows 6
3.1. Unexpected traffic Flows 7
3.1.1. Unexpected traffic flows caused by local filtering of overlapping prefixes 8
3.1.2. Unexpected traffic flows caused by remotely triggered filtering of overlapping prefixes 12
4. Techniques to detect unexpected traffic flows caused by filtering of overlapping prefixes 15
4.1. Being the 'victim' of unexpected traffic flows 15
4.2. Being a contributor to the existence of unexpected traffic flows in other networks 15
5. Techniques to counter unexpected traffic flows due to the filtering of overlapping prefixes 16
5.1. Reactive counter-measures 17
5.2. Anticipant counter-measures 18
5.2.1. Access lists 18
5.2.2. Automatic overlapping prefix filtering 19
5.2.3. Neighbor-specific forwarding 19
6. Conclusions 19
7. References 20
7.1. References 0
7.2. URIs 20
Authors' Addresses 20

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 can 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 [1]). 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 an Autonomous System filters overlapping prefixes and forwards packets according to the covering prefix, the discrepancy in the routing policies applied to covering and overlapping prefixes can create unexpected traffic flows that infringe the policies of other ASes still holding a path towards the overlapping prefix.

This document presents examples of such cases 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 Internet Service Providers (ISPs).

The rest of the document is organized as follows: Section 2 illustrates the motivation to filter overlapping prefixes. In Section 3, we provide some scenarios in which the filtering of overlapping prefixes lead to the creation of unexpected traffic flows on other ASes. Section 4 and Section 5 discuss some techniques that ASes can use for, respectively, detect and react to unexpected traffic flows.

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.

2.1. Local filtering

Let us first analyze the scenario depicted in Figure 1. AS1 and AS2 are two autonomous systems spanning a large geographical area and peering in 3 different physical locations. Let AS1 announce prefix 10.0.0.0/22 over all peering links with AS1. 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 destined to prefix 10.0.0.0/24 on the link closer to this subnet, 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 legitimately detect and filter prefix 10.0.0.0/24.

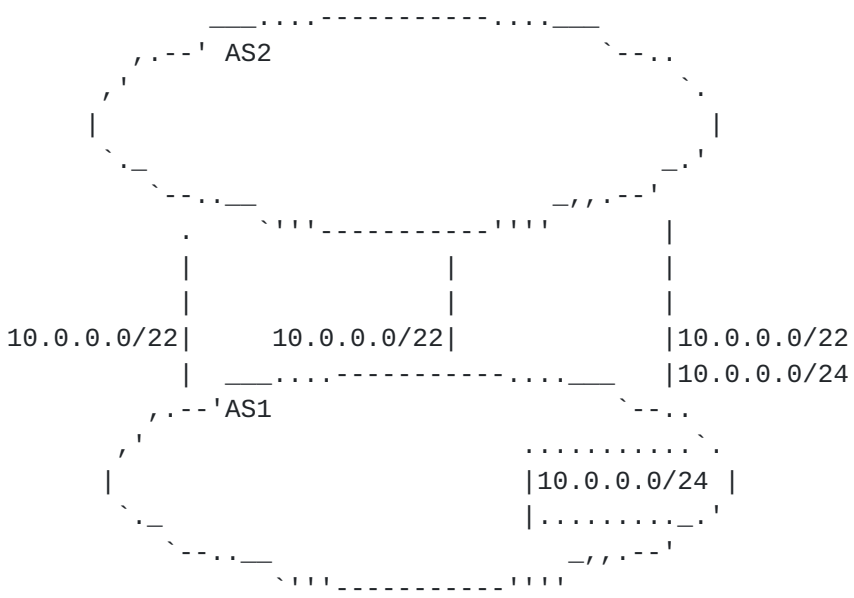

Figure 1: Basic scenario of local filtering

Local filtering could be required in other cases. 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.

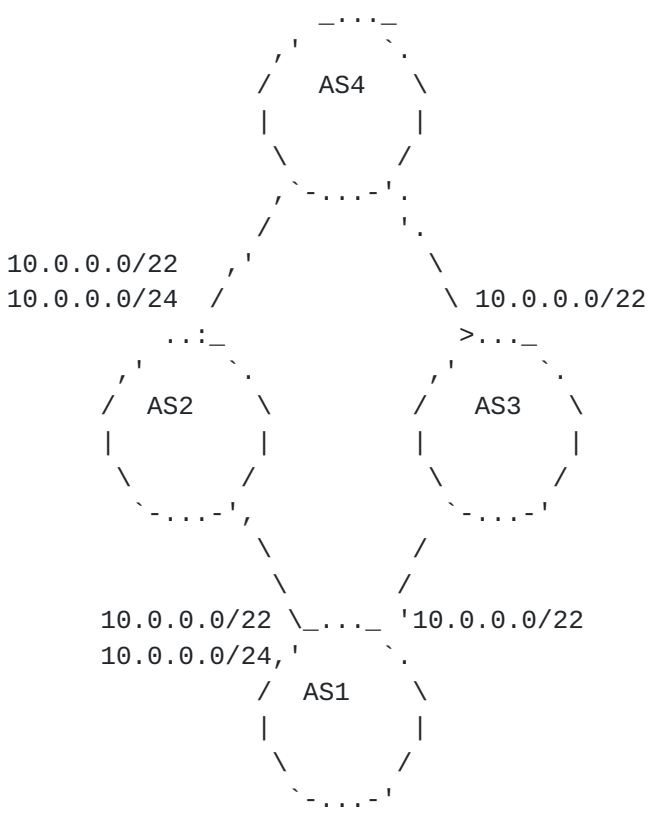

Figure 2: Basic scenario of local filtering

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 prefixes. 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 [2].

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.

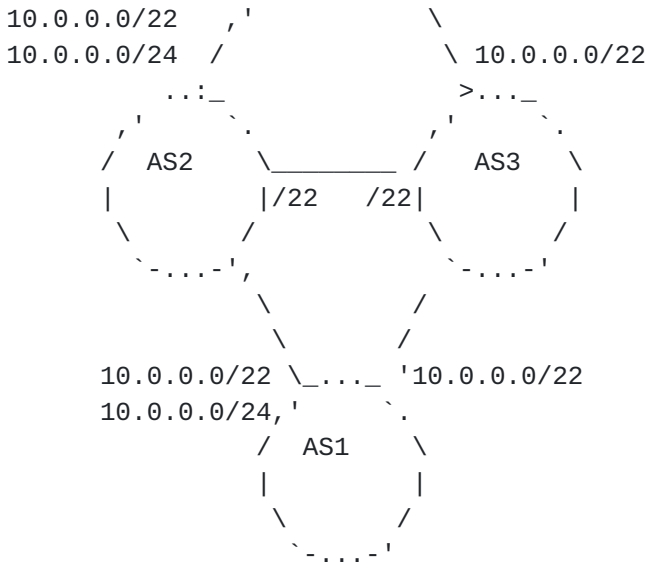

Figure 3: Remote triggered filtering

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 unexpected traffic flows at other, distant, ASes.

3. Uses of overlapping prefix filtering that create unexpected traffic flows

In this section, we define the concept of unexpected traffic flows and describe three configuration scenarios that lead to their creation. Note that these examples do not capture all the cases where such issues can take place.

3.1. Unexpected traffic Flows

The BGP policy of an Internet Service provider includes all actions performed over its originated routes and the routes received externally. One important part of the BGP policy is the selection of the routes that are propagated to each neighboring AS. One of the goals of these policies is to allow ISPs to avoid transporting traffic between two ASes without economical gain. For instance, ISPs typically propagate to their peers only routes coming from its customers (RFC4384 [3]). We briefly illustrate this operation in Figure 4. In the figure, AS2 is establishing a settlement free peering with AS1 and AS3. AS2 receives prefix P3/p3, from AS3. AS2, however, is not interested in transporting traffic from AS1 to AS3, therefore it does not propagate the prefix to AS1. In the figure, we also show a customer of AS2, AS4, which is announcing prefix P4/p4. AS2 propagates this prefix to AS1.

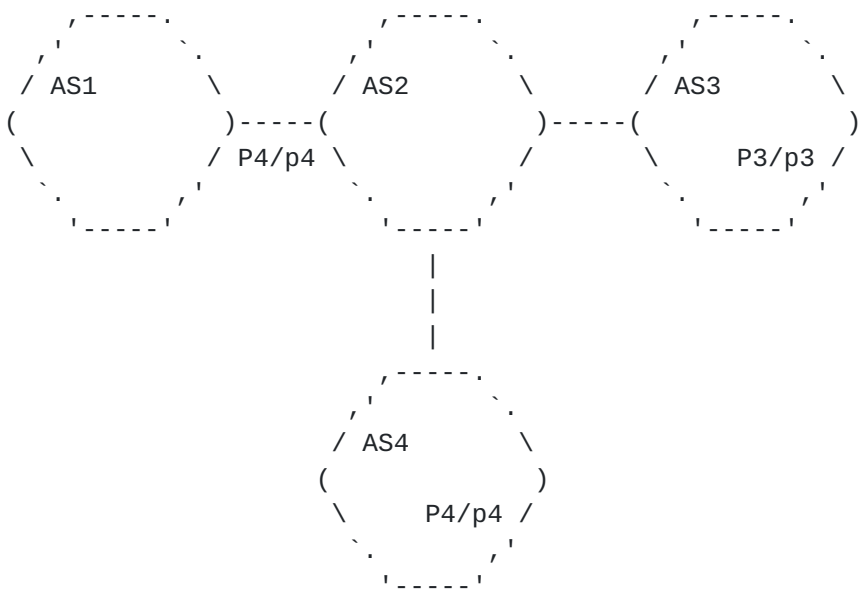

Figure 4: Prefix exchange among four autonomous systems

Although ISPs usually implement the aforementioned policies, unexpected traffic flows may still appear. In Figure 4, unexpected traffic flows are created, when, despite AS2's policy, traffic arriving from peer AS1 is received and transported to AS3 by AS2. These types of traffic flows can arise due to a number of reasons. Specifically, in this document we explain how the filtering of overlapping prefixes might cause unexpected traffic flows on ASes. We provide examples of these cases in the next sections.

3.1.1. Unexpected traffic flows caused by local filtering of overlapping prefixes

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 unexpected traffic flows.

3.1.1.1. Initial setup

We start by describing the basic scenario of this case in Figure 5.

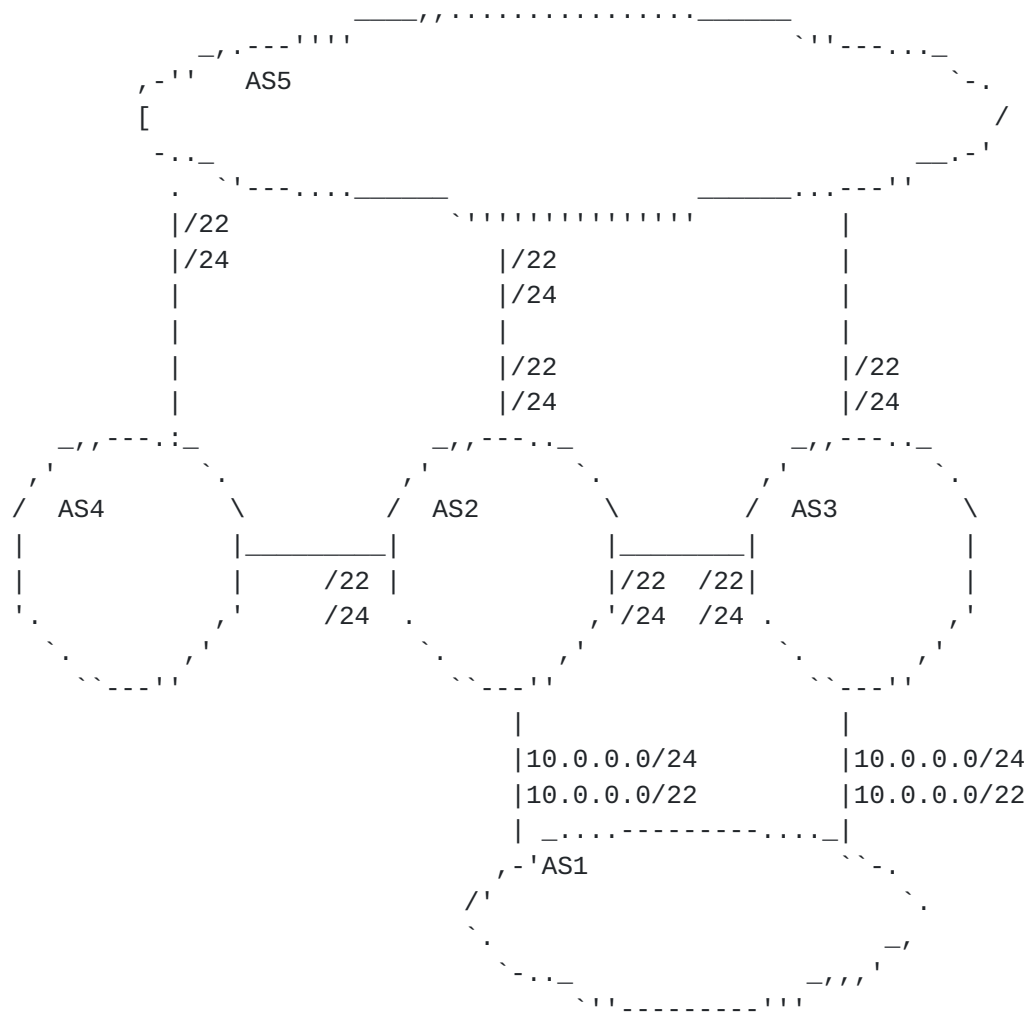

Figure 5: Initial Setup Local

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 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 as best path to AS1 the one received from AS3.

3.1.1.2. Unexpected traffic flows by local filtering - Case 1

In the next scenarios, we show that if AS4 filters the incoming overlapping prefix from AS5, there is a situation in which unexpected traffic flows are created on other ASes.

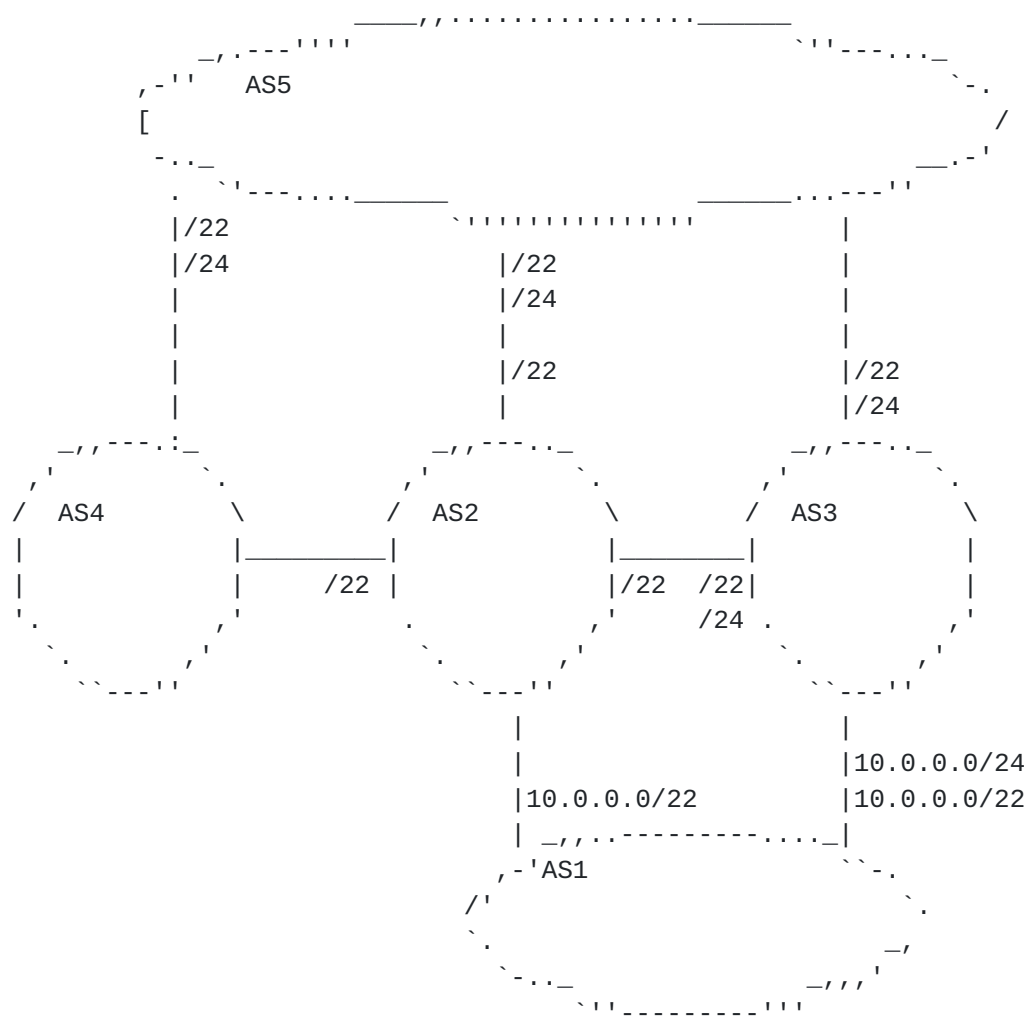

Figure 6: Unexpected traffic flows by local filtering - Case 1

Let us assume the scenario illustrated in Figure 6. For this case, AS1 only propagates the overlapping prefix to AS3. AS4 receives the overlapping prefix only from its transit provider, AS5.

AS4 now is 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 to prefix 10.0.0.0/24 is forwarded towards AS2. Because AS2 receives the more specific prefix from AS3, traffic from AS4 to prefix 10.0.0.0/24 follows the path AS4-AS2-AS3-AS1. AS2's BGP policies are implemented to avoid using itself to exchange traffic between AS4 and AS3. However, due to the discrepancies of routes from the overlapping and covering prefixes, unexpected traffic flows between AS4 and AS3 still exist on AS2's network. This situation is economically detrimental for AS2, since it forwards traffic from a peer to a non-customer neighbor.

3.1.1.3. Unexpected traffic flows by local filtering - Case 2

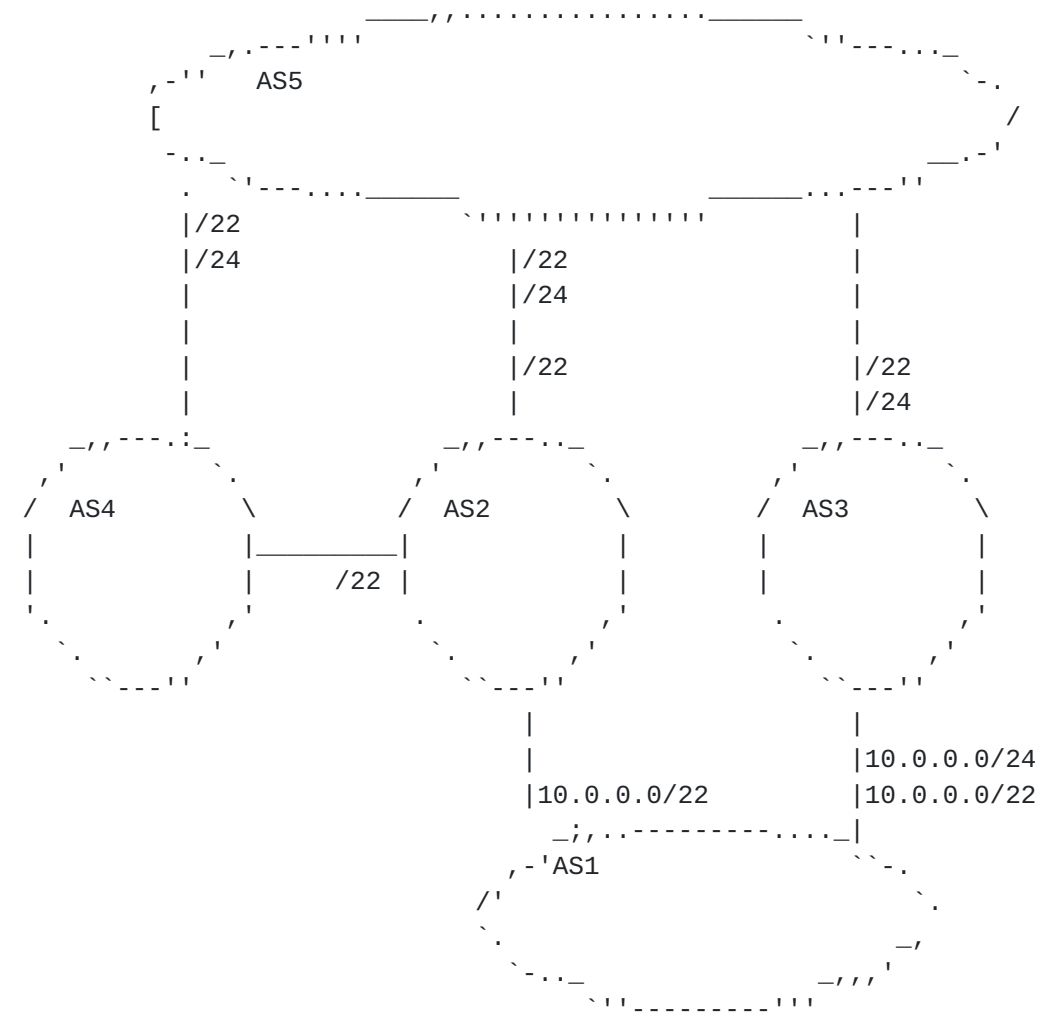

Figure 7: Unexpected traffic flows after local filtering - Case 2

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 7.

Similar to the scenario described in Section 3.1.1.2, 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 would be forwarded towards AS2. Due to the existence of a route to prefix 10.0.0.0/24, AS2 receives the traffic heading to this prefix from AS4 and sends it to AS5. This situation creates unexpected traffic flows that contradict AS2's BGP policy,

since the AS ends up forwarding traffic from a peer to a transit network.

3.1.2. Unexpected traffic flows caused by remotely triggered filtering of overlapping prefixes

We present a configuration scenario in which an AS, using the mechanism described in Section 2.2, informs its provider to selectively propagate an overlapping prefix, leading to the creation of unexpected traffic flows in another AS.

3.1.2.1. 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 A1's path as best, and propagate it 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.

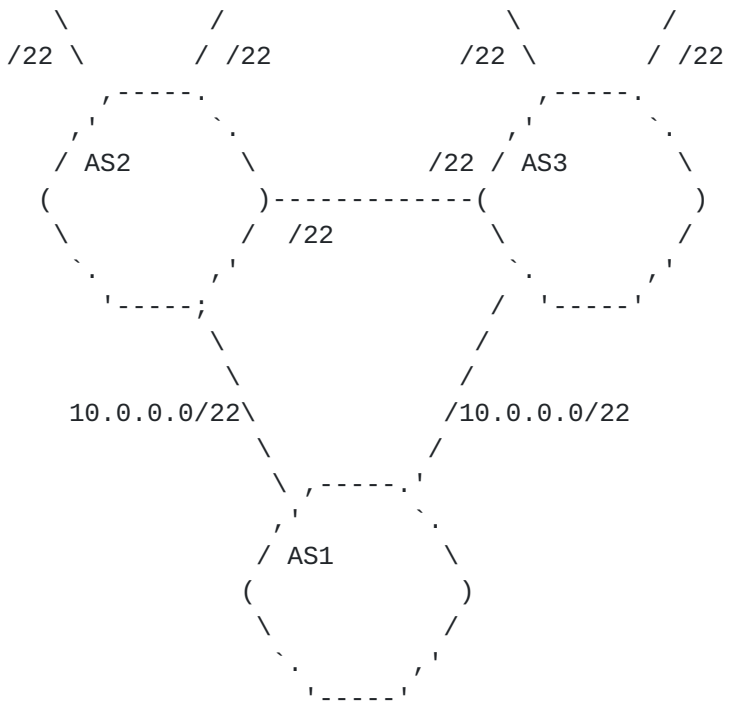

Figure 8: Example scenario

3.1.2.2. 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.

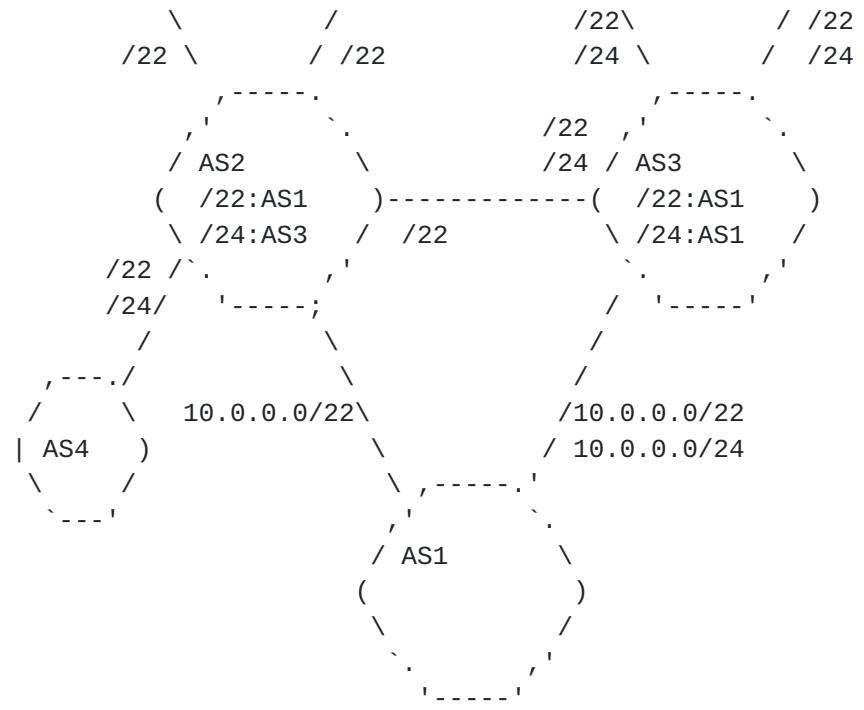

Figure 9: Injection of overlapping prefix

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.

3.1.2.3. Creation of unexpected traffic flows 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.

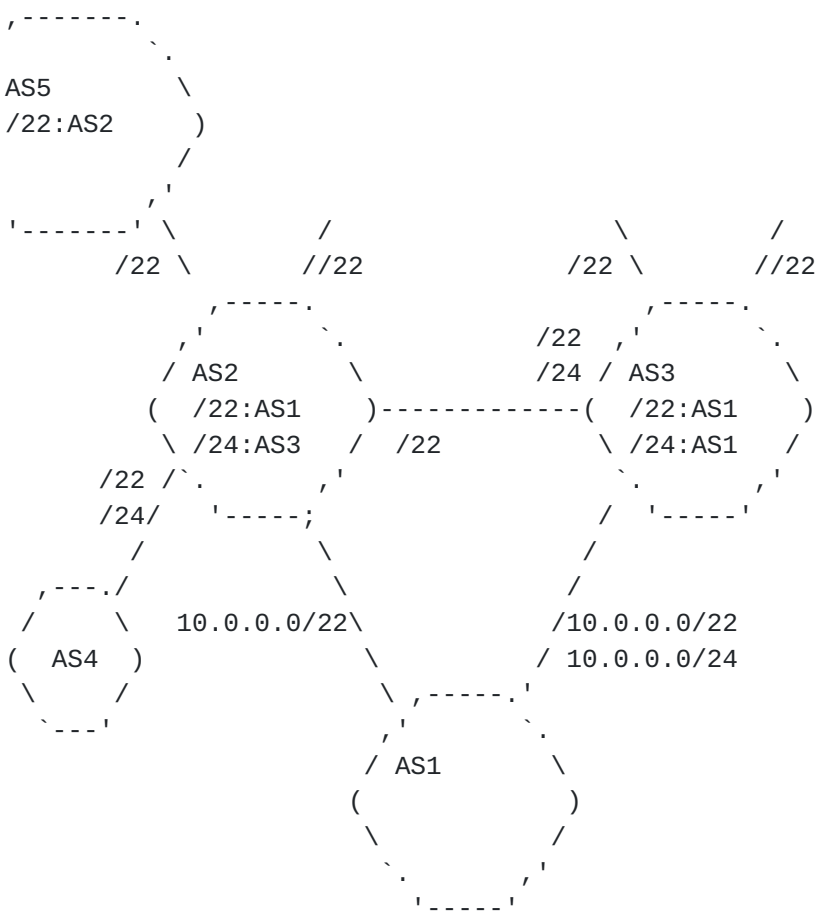

Figure 10: More Specific Injection

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

ASes that are not customers of AS2 will not receive a path for 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 in which traffic flows between non-customer ASes take place in AS2.

4. Techniques to detect unexpected traffic flows caused by filtering of overlapping prefixes

We differentiate the techniques available for detecting unexpected traffic flows caused by the described scenarios from the cases in which the interested AS is the victim or contributor of such operations.

4.1. Being the 'victim' of unexpected traffic flows

To detect if unexpected traffic flows are taking place in its network, an ISP can monitor its traffic data and validate if any flow entering the ISP network through a non-customer link is forwarded to a non-customer next-hop.

As mentioned in Section 3.1, unexpected traffic flows might appear due to different situations. To discover if the problem arose after the filtering of prefixes by neighboring ASes, an operator can analyze available BGP data. For instance, 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 and not the path towards the overlapping prefix. This situation 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 existence of unexpected traffic flows in other networks

It can be considered problematic to be causing unexpected traffic flows on other ASes. This situation may appear 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 the creation of unexpected traffic flows is not a practical option.

Traffic data does not help an ISP detect that it is acting as a contributor of the creation of the unexpected traffic flow. 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 compared with their terms of use.

5. Techniques to counter unexpected traffic flows due to the filtering of overlapping prefixes

Network Operators can adopt different approaches with respect to unexpected traffic flows. 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 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 configuration scenario is set up.

We use the scenario depicted in Figure 11 to describe these two kinds of approaches. Based on our analysis, we observe that anticipant approaches can be complex to implement and can lead to undesired repercussions. Therefore, we conclude that the reactive approach is the more reasonable recommendation to deal with unexpected flows.

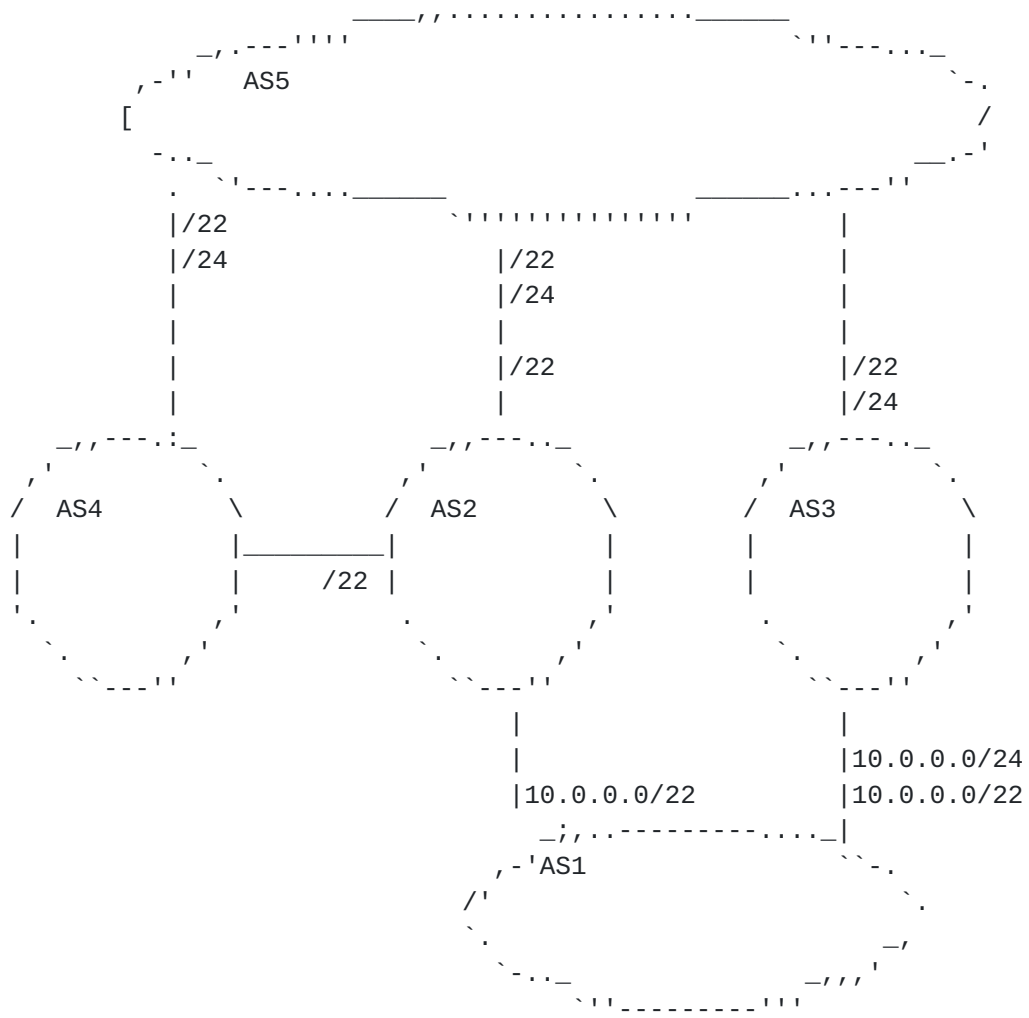

Figure 11: Anticipant counter-measures - Base example

5.1. Reactive counter-measures

An operator who detects unexpected traffic flows originated by any of the cases described in Section 3 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 overlapping 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 covering prefix at the peering session with the neighboring AS from which it is receiving traffic to the overlapping prefix. In the example of Figure 11, AS2 would filter out the prefix 10.0.0.0/22 from the eBGP session with AS4. In this case, all the traffic heading to the prefix 10.0.0.0/22 from AS1 would not longer traverse AS2. AS2 should evaluate if solving the inconvenient originated by the unexpected traffic flows are worth the loss of this traffic share.

- 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 11, 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. However, as AS2 will no longer possess a route to the overlapping prefix, it risks losing the traffic share from customers different from AS1 to that prefix. Furthermore, this action can generate conflicts between AS2 and AS1, since AS2 does not follow the policy expressed by AS1 in its BGP announcements.

It is possible that the behavior from the neighboring AS that is causing the unexpected traffic flows opposes the peering agreement. In this case, an operator can account the amount of traffic that has been subject to the unexpected flows 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.

5.2. Anticipant counter-measures

5.2.1. Access lists

An operator can configure its routers to install dynamically an access-list made of the prefixes towards which the forwarding of traffic from that interface would lead to unexpected traffic flows. In the example of Figure 11, 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 valid route towards 10.0.0.0/22.

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 policy conflicts and hence had no means to avoid the creation of unexpected traffic flows.

5.2.2. Automatic overlapping prefix filtering

As described in Section 3, filtering of overlapping prefixes can in some scenarios lead to unexpected traffic flows. 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 11: if AS2 or AS3 filter prefix 10.0.0.0/24, there would be no unexpected traffic flow in AS2. Nevertheless, as described in Section 5.1, the filtering of overlapping prefixes can generate conflicts between AS1 and AS2, since AS2 would not forward traffic according to AS1's policy. Additionally, AS2 can lose traffic share for the overlapping prefix from customers different from AS1.

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 11 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 of 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 unexpected traffic flow would occur.

Different techniques could provide the functionality just described; however, their technical implementation can be complex to design and operate. [2] describes an approach to implement this behavior. Similar to the solution described in Section 5.2.2, this approach could create conflicts between AS2 and AS1, since the traffic forwarding performed by A2 goes against the policy of AS1.

6. Conclusions

In this document, we described threats to policies of autonomous systems caused by the filtering of overlapping prefixes performed by external networks. We provide 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 potential victims 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 overlapping prefixes, we encourage that network operators implement this type of filters only after considering the cases described in this document.

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>.

7.2. URIs

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

[2] http://tools.ietf.org/html/draft-white-grow-overlapping-routes-02

[3] http://www.ietf.org/rfc/rfc4384.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