Network Working Group
Internet-Draft
Intended status: Best Current Practice
Expires: December 31, 2015

S. Vinapamula
Juniper Networks
M. Boucadair
France Telecom

June 29, 2015

Recommendations for Prefix Binding in the Softwire DS-Lite Context
draft-vinapamula-softwire-dslite-prefix-binding-06

Abstract

This document discusses issues induced by the change of the Dual-Stack Lite (DS-Lite) Basic Bridging BroadBand (B4) IPv6 address and sketches a set of recommendations to solve those issues.

Status of This Memo

Copyright Notice

Table of Contents

1. Introduction

IPv6 deployment models assume IPv6 prefixes are delegated by Service Providers to the connected CPEs (Customer Premises Equipments) or hosts, which in turn derive IPv6 addresses out of that prefix. In the case of DS-Lite [RFC6333], that is an IPv4 service continuity mechanism over an IPv6 network, the Basic Bridging BroadBand (B4) element derives an IPv6 address for the IPv4-in-IPv6 softwire setup purposes.

The B4 element might obtain a new IPv6 address, for a variety of reasons that include (but are not limited to) a reboot of the CPE, power outage, DHCPv6 lease expiry, or other actions undertaken by the Service Provider. If this occurs, traffic forwarded to a B4's previous IPv6 address may never reach its destination or be delivered to another B4 that now uses the address formerly assigned to the original B4. This situation affects all mapping types, both implicit (e.g., by sending a TCP SYN) and explicit (e.g., using Port Control Protocol (PCP) [RFC6887]). The problem is further elaborated in Section 2.

This document proposes recommendations to soften the impact of such renumbering issues (Section 4).

Note that in some deployments, CPE renumbering may be required to accommodate some privacy-related requirements to avoid the same prefix be assigned to the same customer. It is out of scope of this document to discuss such contexts.

This document complements [RFC6908].

1.1. Requirements Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in [RFC2119].

2. The Problem

Since private IPv4 addresses assigned to hosts serviced by a B4 element are overlapping across multiple CPEs, the IPv6 address of a B4 element plays a key role in de-multiplexing connections, enforcing policies, and in identifying associated resources assigned for each of the connections maintained by the Address Family Transition Router (AFTR, [RFC6333]). For example, these resources maintain state of Endpoint-Independent Mapping (EIM, Section 4.1 of [RFC4787]), Endpoint-Independent Filtering (EIF, Section 5 of [RFC4787]), preserve the external IPv4 address assigned in the AFTR (i.e., "IP address pooling" behavior as defined in Section 4.1 of [RFC4787]), PCP mappings, etc.

However, the IPv6 address used by the B4 element may change for some reason, e.g., because of a change in the CPE itself or may be because of privacy extensions enabled for generating the IPv6 address (e.g., [RFC7217] or [RFC4941]). Whenever the B4's IPv6 address changes, the associated mappings created in the AFTR are no more valid. This may result in the creation of a new set of mappings in the AFTR.

Furthermore, a misbehaving user may be tempted to change the B4's IPv6 address in order to "grab" more ports and resources at the AFTR side. This behavior can be seen as a potential Denial of Service (DoS) attack from misbehaving users. Note that this DoS attack can be achieved whatever the port assignment policy enforced by the AFTR (individual ports, port sets, randomized port bulks, etc.).

Service Providers may want to enforce policies in order to limit the usage of the AFTR resources on a per-subscriber basis for fairness resource usage (see REQ-4 of [RFC6888]). These policies are used for dimensioning purposes and also to ensure that AFTR resources are not exhausted. To that aim, a subscriber should be identified by the AFTR based upon the IPv6 prefix assigned to the corresponding CPE, and not according to the derived B4's IPv6 address. Also, whenever the B4's IPv6 address, enforcing policies based on this address doesn't resolve stale mappings hanging around in the system, consuming not only system resources, but also reducing the available quota of resources per subscriber. Clearing those mappings can be envisaged, but that will cause a lot of churn in the AFTR and could be disruptive to existing connections, which is not desirable.

When application servers are hosted behind a B4 element, and when there is a change of the B4's IPv6 address which results in a change of the external IPv4 address and/or the external port number at the AFTR side, these servers have to advertise about their change (see Section 1.1 of [RFC7393]). Means to discover the change of B4's IPv6 address, the external IPv4 address and/or the external port are therefore required. Latency issues are likely to be experienced when an application server has to advertise its newly assigned external IPv4 address and port, and the application clients have to discover that newly assigned address and/or port and re-initiate connections with the application server.

A solution to these problems is to enforce policies based on the IPv6 prefix assigned to DS-Lite serviced subscribers instead of the B4's IPv6 address. Section 3 introduces the subscriber-mask that is meant to derive the IPv6 prefix assigned to a subscriber's CPE from the source IPv6 address of a packet received from a B4 element.

3. Introducing Subscriber-Mask

The subscriber-mask is defined as an integer that indicates the length of significant bits to be applied on the source IPv6 address (internal side) to identify unambiguously a CPE.

Subscriber-mask is an AFTR system-wide configuration parameter that is used to enforce generic per-subscriber policies. Applying these generic policies does not require to configure every subscriber's prefix.

Subscriber-mask must be configurable; the default value is 56.

Example: suppose the 2001:db8:100:100::/56 prefix is assigned to a DS-Lite enabled CPE. Suppose also that the 2001:db8:100:100::1 address is the IPv6 address used by the B4 element that resides in that CPE. When the AFTR receives a packet from this B4 element (i.e., the source address of the IPv4-in-IPv6 packet is 2001:db8:100:100::1), the AFTR applies the subscriber-mask (e.g., 56) on the source IPv6 address to compute the associated prefix for this B4 element (that is 2001:db8:100:100::/56). Then, the AFTR enforces policies based on that prefix (2001:db8:100:100::/56), not on the exact source IPv6 address.

4. Recommendations

In order to mitigate the issues discussed in Section 2, the following recommendations are made:

1. A policy SHOULD be enforced at the AFTR to limit the number of active DS-Lite softwires per subscriber. The default value MUST be 1.

   This policy aims to prevent a misbehaving subscriber to mount several DS-Lite softwires that would consume additional AFTR resources (e.g., get more external ports if the quota was enforced on a per-softwire basis, consume extra processing induced by a large number of active softwires).

2. Resource contexts created and maintained by the AFTR SHOULD be based on the delegated IPv6 prefix instead of the B4's IPv6 address. The AFTR derives the delegated prefix from the B4's IPv6 address by means of configured subscriber-mask (Section 3). Administrators SHOULD configure per-prefix limits of resource usage, instead of per-tunnel limits. These resources include the maximum number of active flows, the maximum number of PCP-created mappings, NAT pool resources, etc.

3. In the event a new IPv6 address is assigned to the B4 element, the AFTR SHOULD migrate existing state to be bound to the new IPv6 address. This operation ensures that traffic destined to the previous B4's IPv6 address will be redirected to the newer B4's IPv6 address. The destination IPv6 address for tunneling return traffic from the AFTR SHOULD be the last seen as B4's IPv6 source address from the CPE.

   This recommendation avoids stale mappings at the AFTR and minimizes the risk of service disruption for subscribers.

   The AFTR uses the subscriber-mask to determine whether two IPv6 addresses belong to the same CPE (e.g., if the subscriber-mask is set to 56, the AFTR concludes that 2001:db8:100:100::1 and 2001:db8:100:100::2 belong to the same CPE assigned with 2001:db8:100:100::/56).

4. In the event of change of the CPE WAN's IPv6 prefix, unsolicited PCP ANNOUNCE messages SHOULD be sent by the B4 element to internal hosts connected to the PCP-capable CPE so that they update their mappings accordingly.

   This allows internal PCP clients to update their mappings with the new B4's IPv6 address and to trigger updates to rendez-vous servers (e.g., dynamic DNS). A PCP-based dynamic DNS solution is specified in [RFC7393].

5. When a new prefix is assigned to the CPE, stale mappings may exist in the AFTR. This will consume both implicit and explicit

resources. In order to avoid such issues, stable IPv6 prefix assignment is RECOMMENDED.

6. In case for any reason an IPv6 prefix has to be reassigned, it is RECOMMENDED to reassign an IPv6 prefix (that was previously assigned to a given CPE) to another CPE only when all the resources in use associated with that prefix are cleared from the AFTR. Doing so avoids to redirect traffic, destined to the previous prefix owner, to the new one.

5. Security Considerations

Security considerations related to DS-Lite are discussed in [RFC6333].

Enforcing the recommendations documented in Section 4 together with rate limiting softwires with new source IPv6 addresses from the same CPE defend against DoS attacks that would result in varying the B4's IPv6 address to exhaust AFTR resources. A misbehaving CPE can be blacklisted by enforcing appropriate policies based on the prefix derived from the subscriber-mask.

6. IANA Considerations

This document does not require any action from IANA.

7. Acknowledgements

G. Krishna, C. Jacquenet, I. Farrer, Y. Lee, Q. Sun, and R. Weber provided useful comments. Many thanks to them.

8. References

8.1. Normative references

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.

[RFC6333] Durand, A., Droms, R., Woodyatt, J., and Y. Lee, "Dual-Stack Lite Broadband Deployments Following IPv4 Exhaustion", RFC 6333, August 2011.

[RFC6887] Wing, D., Cheshire, S., Boucadair, M., Penno, R., and P. Selkirk, "Port Control Protocol (PCP)", RFC 6887, April 2013.

8.2. Informative references

[RFC4787] Audet, F. and C. Jennings, "Network Address Translation (NAT) Behavioral Requirements for Unicast UDP", BCP 127, RFC 4787, January 2007.

[RFC4941] Narten, T., Draves, R., and S. Krishnan, "Privacy Extensions for Stateless Address Autoconfiguration in IPv6", RFC 4941, September 2007.

[RFC6888] Perreault, S., Yamagata, I., Miyakawa, S., Nakagawa, A., and H. Ashida, "Common Requirements for Carrier-Grade NATs (CGNs)", BCP 127, RFC 6888, April 2013.

[RFC6908] Lee, Y., Maglione, R., Williams, C., Jacquenet, C., and M. Boucadair, "Deployment Considerations for Dual-Stack Lite", RFC 6908, March 2013.

[RFC7217] Gont, F., "A Method for Generating Semantically Opaque Interface Identifiers with IPv6 Stateless Address Autoconfiguration (SLAAC)", RFC 7217, April 2014.

[RFC7393] Deng, X., Boucadair, M., Zhao, Q., Huang, J., and C. Zhou, "Using the Port Control Protocol (PCP) to Update Dynamic DNS", RFC 7393, November 2014.

Authors' Addresses

Suresh Vinapamula
Juniper Networks
1194 North Mathilda Avenue
Sunnyvale, CA 94089
USA

Phone: +1 408 936 5441
EMail: sureshk@juniper.net

Mohamed Boucadair
France Telecom
Rennes 35000
France

EMail: mohamed.boucadair@orange.com