SAVI
Internet Draft
Intended status: Standard Tracks
Expires: May 2010

J. Bi, J. Wu, G. Yao
CERNET
F. Baker
Cisco

November 9, 2009

SAVI Solution for DHCPv4/v6
draft-bi-savi-dhcp-00.txt

Status of this Memo

Abstract

This document specifies the procedure for creating bindings between a DHCPv4 [RFC2131]/DHCPv6 [RFC3315] assigned source IP address and an anchor (refer to [SAVI-framework]) on SAVI (Source Address Validation Improvements) device. The bindings can be used to filter packets with forged IP addresses generated on the local link.

Table of Contents

1. Introduction

This document describes the procedure for creating bindings between DHCP assigned addresses and an anchor (refer to [savi-framework]). Other related details about this procedure are also specified in this document.

These bindings can be used to filter packets with forged IP addresses. How to use these bindings is specified in [savi-framework], depending on the environment and configuration. The definition and examples of anchor is also specified in [savi-framework].

The binding process is inspired by the work of IP source guard. This specification differs from IP source guard in the specification for

collision detection, which is quite useful in an environment with multiple address assignment methods.

2. Conventions used in this document

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].

3. Mechanism Overview

The mechanism specified in this document is designed to provide a host level source IP address validation granularity, as a supplement to BCP38 [BCP38]. This mechanism is deployed on the access device (including access switch, wireless access point/controller, etc), and performs mainly DHCPv4/v6 snooping to set up bindings between DHCP assigned IP address and corresponding anchors. The bindings can be used to validate the source address in the packets.

4. Background and Related Protocols

This mechanism is an instance of a SAVI [savi-framework] solution, specialized for addresses assigned using the DHCP protocol.

Dynamic Host Configuration Protocol version 4 [RFC2131] and Dynamic Host Configuration Protocol version 6 [RFC3315] specify the procedures for providing a client with assigned address and other configuration information from a DHCP server. If a client gets an address through the DHCP protocol, the address should be regarded as a potential "authorized" or "registered" address of the client.

In IPv6, IPv6 Stateless Autoconfiguration [RFC4862] is used as another address assignment mechanism. A node can use this mechanism to auto-configure an IPv6 address. A DHCPv6 client may use a stateless address to send message to DHCP server. Even in a DHCPv6-only environment, a node must assign its link-local address through this mechanism. [RFC4862] also clearly requires that duplicated address detection must be performed on any IPv6 address, including DHCPv6 address.

[RFC4861] specifies the Neighbor Discovery protocol, which is an essential part of IPv6 address assignment.

[RFC5227] specifies the procedure to detect IPv4 address collision. It is not required currently. However, this feature is useful to determine the uniqueness of an IPv4 address on the link.

5. Terminology

The terms used in this document are described in [savi-framework], [RFC2131] and [RFC3315].

6. Conceptual Data Structures

(To be removed and merged with data structures used by other mechanisms in [savi-framework] if possible)

This section describes the possible conceptual data structures used in this mechanism.

Two main data structures are used to record bindings and their states respectively. There is redundancy between the two structures, for the consideration of separation of data plane and control plane.

6.1. Binding State Table (BST)

This table contains the state of binding between source address and anchor. Entries are keyed on the anchor and source IP address. Each entry has a lifetime field recording the remaining lifetime of the entry, a state field recording the state of the binding and a field for recording other information.

Anchor	Address	State	Lifetime	Other
A	IP_1	Bound	65535	
A	IP_2	Bound	10000	
B	IP_3	_Start	1	

Figure 1 Instance of BST

6.2. Filtering Table (FT)

This table contains the bindings between anchor and address, keyed on anchor. This table doesn't contain any state of the binding. This table is only used to filter packets. An Access Control List can be regarded as a practical instance of this table.

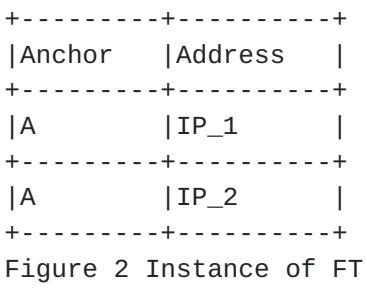

Figure 2 Instance of FT

7. Binding States Description

This section describes the binding states of this mechanism.

START A DHCP request (or a DHCPv6 Confirm) has been received from host, and it may trigger a new binding.

LIVE A DHCP address has been acknowledged by a DHCP server.

DETECTION A gratuitous ARP or Duplicate Address Detection NSOL has been sent by the host (or the SAVI device).

BOUND The address has passed duplicate detection and it is bound with the anchor.

8. DHCP Scenario

(This section should be removed and merged with other scenarios in [savi-framework])

This section specifies the deployment scenarios of this mechanism.

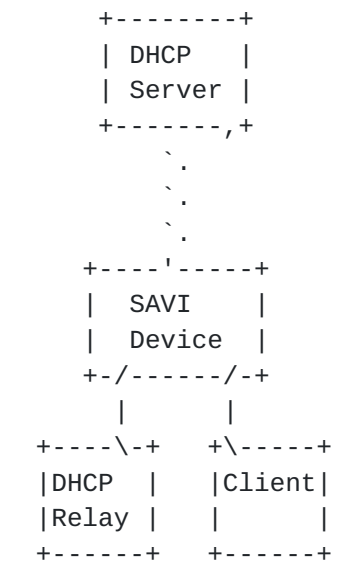

Figure 3 DHCP Scenario

9. Anchor Attributes

This section specifies the anchor attributes involved in this mechanism.

9.1. SAVI-Host Attribute

(This attribute should be described in the [savi-framework])

If and only if an anchor is exclusively associated with a single host, this anchor can be set to have SAVI-Host attribute.

9.2. SAVI-Poly Attribute

(This attribute should be described in the [savi-framework])

If an anchor is associated with a small number of hosts through a converged device, this anchor can be set to have the SAVI-Poly attribute. The SAVI-Poly attribute is mutually exclusive with the SAVI-Host attribute.

The main difference in process on an anchor with the SAVI-Poly attribute from one with the SAVI-Host attribute is the policy of binding consistency.

9.3. SAVI-DHCP-Trust Attribute

If and only if an anchor is associated with a trustable DHCP server/relay, it can be set to have this attribute.

If DHCP is used to assign address in the network, there MUST be at least one anchor with this attribute. DHCP Reply message not coming from such ports MUST be dropped.

9.4. SAVI-RA-Trust Attribute

(This attribute should be described in the [savi-framework])

If and only if an anchor is associated with a trustable router, it can be set to have this attribute.

There MAY be no SAVI-RA-Trust anchor on a SAVI device.

Router Advertisement not received from a SAVI-RA-Trust anchor MUST be discarded.

9.5. SAVI-SAVI Attribute

(This attribute should be described in the [savi-framework])

If and only if an anchor is associated with another SAVI device, it can be set to have this attribute.

This attribute is mutually exclusive with SAVI-Host and SAVI-Poly attribute.

10. Prefix Configuration

(This section should be included in [SAVI-framework] but not this document.)

Before setting up a host-level granularity binding table, it is important to configure correct prefixes on the SAVI device. At least two prefix scopes must be set: the IPv4 prefix and IPv6 prefixes. This document suggests set 3 prefix scopes:

IPv4 Prefix:

The allowed scope of any kind of IPv4 addresses. It can be set manually.

IPv6 SLAAC Prefixes:

The allowed scope of SLAAC and manually configured IPv6 addresses. It can be set through snooping RA message from port with SAVI-RA-Trust attribute, or manual configuration.

FE80::/64 MUST be set to a feasible prefix.

IPv6 DHCPv6 Prefixes:

The allowed scope of DHCPv6 addresses. It can be set through RA snooping, DHCP-PD protocol, or manual configuration.

If some of the prefix scope is set to have non prefix, it implies corresponding address assignment method is not allowed in the network.

There is no need to explicitly present these prefix scopes. But these restrictions MUST be used as premier check in binding set up.

Refer to security consideration for other discussions.

11. Binding Set Up

This section specifies the procedure of setting up bindings based on control packet snooping.

11.1. Process of DHCP Snooping

11.1.1. Initialization

This procedure will not be performed if:

1. Option 82 is used to keep anchor in DHCP Request and Reply, or

2. Unspoofable MAC is used as anchor(802.11i,802.1ae/af), or

3. The mapping table from MAC to anchor is secure.

If none of these three requirements are satisfied, this procedure MUST be performed.

11.1.1.1. Trigger Event

A DHCPv4/v6 Request or a DHCPv6 Confirm is received with an anchor which has the attribute of SAVI-Host or SAVI-Poly.

11.1.1.2. Message Validation

The SAVI device checks the Request or Confirm as follows:

1. Whether the limitation on binding entry number of this anchor will be exceeded.

11.1.1.3. Following Actions

Forward the REQUEST message if binding entry limitation will not be exceeded.

Generate an entry in the Binding State Table (BST) and set the state field to START. The lifetime of this entry is set to be MAX_DHCP_RESPONSE_TIME. The Transaction ID (Refer to Section 2 in [RFC2131] and Section 4.2 in [RFC3315]) field of the request packet is also recorded in the entry.

Anchor	Address	State	Lifetime	Other
A		Start	MAX_DHCP_RESPONSE_TIME	TID

Figure 4 Binding entry in BST on initialization

The TID is kept for assurance that the response from the DHCP server can be delivered to the request host. This is left as an open issue for future discussion.

11.1.2. From START to LIVE

11.1.2.1. Trigger Event

A DHCPv4 DHCPACK or DHCPv6 REPLY message is received.

11.1.2.2. Message Validation

The SAVI device checks the message as follows:

1. Whether the message is received with an anchor which has the SAVI-DHCP-Trust attribute;

2. Whether the entry in the BST with corresponding TID is in the START state.

11.1.2.3. Following Actions

Deliver the message to the destination.

Set the state of the corresponding entry to be LIVE. The lifetime of the entry is set to be MAX_ARP_PREPARE_DELAY or MAX_DAD_PREPARE_DELAY respectively. The lease time is also recorded in the entry.

Anchor	Address	State	Lifetime	Other
A	Addr	LIVE	MAX_ARP_PREPARE_DELAY or MAX_DAD_PREPARE_DELAY	Lease Time

Figure 5 Binding entry in BST on assignment

Or set the state of the corresponding entry to be DETECTION, and send an ARP Request or NSOL for the assigned address.

Anchor	Address	State	Lifetime	Other
A	Addr	DETECTION	MAX_ARP_DELAY or MAX_DAD_DELAY	Lease Time

Figure 6 Binding entry in BST on assignment: another case

Insert an entry into the Filtering Table if the assigned address is not bound with another anchor.

Anchor	Address
A	Addr

Figure 7 Binding entry in FT on assignment

11.1.3. From LIVE to DETECTION

(This section should be removed or modified if all the DAD related procedures are to be described in SLAAC document)

11.1.3.1. Trigger Event

A gratuitous ARP Request or Duplicate Address Detection Neighbor Solicitation is received from the host. Or a timeout event of an entry with state LIVE occurs.

11.1.3.2. Message Validation

The SAVI device checks the message as follows:

1. Whether the Target IP address field of the ARP Request or Neighbor Solicitation has been bound with the corresponding anchor in BST or FT.

11.1.3.3. Following Actions

If timeout event of an entry with state LIVE happens, send an ARP Request or a DAD NSOL, with target address set to the recorded address in the entry.

Set the state of the entry to be DETECTION. The lifetime of the entry is set to be MAX_ARP_DELAY or MAX_DAD_DELAY respectively.

Anchor	Address	State	Lifetime	Other
A	Addr	DETECTION	MAX_ARP_DELAY or MAX_DAD_DELAY	Lease Time

Figure 8 Binding entry in BST on detection

11.1.4. From DETECTION to BOUND

11.1.4.1. Trigger Event

A timeout event of an entry with state DETECTION occurs or an ARP Response or NA for an address in BST with state DETECTION is received.

11.1.4.2. Following Actions

If a timeout event of an entry with state DETECTION occurs, set the state of the entry to be BOUND. The lifetime of the entry is set to be the Lease time acknowledged by DHCP server.

Anchor	Address	State	Lifetime	Other
A	Addr	BOUND	Lease time	

Figure 9 Binding entry in BST on finalization

If an ARP Response or NA for an address in BST with state DETECTION is received, remove the corresponding entry in BST and FT.

11.1.5. After BOUND

Whenever a DHCP Decline is received from the host, delete the entry in BST and FT.

Whenever a DHCP Release is received from the host, if the state of the entry is BOUND, delete the entry in BST and FT.

If a DHCPv4 Acknowledgement or DHCPv6 Reply with Renew/Rebind sign is received from the server, set lifetime of the entry in BST to be the new lease time.

If the lifetime of an entry with state BOUND expires, delete the entry in BST and Filter Table.

11.2. State Machine of DHCP Snooping

State	Packet/Event	Action	Next State
-*	REQUEST/CONFIRM	Set up new entry	START
START	ACK	Record lease time	LIVE
START	Timeout	Remove entry	-
LIVE	Gra ARP REQ/DAD NS	-	DETECTION
LIVE	DECLINE	Remove entry	-
LIVE	Timeout	Send ARP Req/DAD NS	DETECTION
DETECTION	Timeout	-	BOUND
DETECTION	ARP RESPONSE	Remove entry	-
DETECTION	DECLINE	Remove entry	-
BOUND	RELEASE/DECLINE	Remove entry	-
BOUND	Timeout	Remove entry	-
BOUND	RENEW/REBOUND	Set new lifetime	BOUND

*: optional.

12. Filtering Specification

This section specifies how to use bindings to filter packets.

12.1. Filter Data Packet

Data packets with an anchor which has attribute SAVI-Host or SAVI-Poly are filtered. There can be an anchor with neither attribute.

If the source of a packet associated with its anchor is in the FT, this packet will be forwarded; or else the packet MUST be discarded.

12.2. Filter Control Packet

The source address of DHCPv4 Request/Discovery must be set to all zeros.

The source address of DHCPv6 Request/Confirm must be an address associated with the corresponding anchor in FT.

The source address of IPv6 NS and IPv6 gratuitous ARP must pass the check on FT. The source address of DAD NS MUST be unspecified address.

All DHCP Reply/Ack packets MUST be from an anchor with the SAVI-DHCP-Trust attribute.

The target address and source address in all the Neighbor Advertisement packets and ARP replies must also pass the checks on FT, if they are associated with anchors which have SAVI-Host or SAVI-Poly attribute.

13. Format and Delivery of Probe Messages

1. Duplicate detection on behavior of host;

   Message Type: DAD NS, Gratuitous ARP Request

   Format:

   Link layer source - link layer address of host;

   Link layer destination - For IPv6, use multicast address specified in [RFC3307]; For IPv4, use broadcast address;

   IP source - Unspecified address for IPv6; The tentative address for IPv4;

   IP destination - For IPv6, multicast address specified in section 5.4.2 of [RFC4861]; For IPv4, the tentative address;

   Delivery:

   MUST not deliver to the host which the SAVI device is performing DAD on behavior of.

2. Send reply on behavior of host to hold bound address for inactive node;

   Message Type: NA, ARP Response

   Link layer source - link layer address of host;

   Link layer destination - The source address of the detecting node;

   IP source - The target address in the detection message;

   IP destination -The source address of the detecting node;

3. Send probe to detect whether an address is still in use (generally in case of port down/up event).

   Message Type: NUD, ARP Request

   Link layer source - link layer address of the SAVI device;

   Link layer destination - The link layer address of the node;

   IP source - The manage IP address of the SAVI device;

   IP destination - The address is suspicious to be inactive.

14. Data packet trigger on SAVI-Poly anchor

To handle the case of movement from one SAVI-Poly port to another, data packet based binding MUST be performed on SAVI-Poly anchor.

Whenever a packet with source address not bound locally is received from a SAVI-Poly anchor, the SAVI device MUST send a DHCP CONFIRM to the DHCP server to ensure the address can be used on the link; if the address can be used, the SAVI device MUST send a DAD NS on behavior of the host to check the uniqueness of the address. If the address passes these two checks, it can be bound with the anchor.

Data packet triggered binging will cause heavy burden and unpredictable danger on the SAVI device. It is not suggested to configure SAVI-Poly anchor and allow mobility between SAVI-Poly anchors. If this function is turned off, the administrator must be aware of the packets from SAVI-Poly anchor may be filtered incorrectly after movement.

15. Binding Remove

If the lifetime of an entry with state BOUND expires, the entry MUST be removed.

When the SAVI device receives a DAD NS/Gra ARP request target at an address bound and there is no reply from the anchor, if the anchor is a SAVI-Host anchor, hold the binding entry through sending NA/ARP Reply; if the anchor is a SAVI-Poly anchor, remove this binding or hold it.

16. Handle port DOWN event

Whenever a port with attribute SAVI-Poly or SAVI-Host turns down, the bindings with the anchor MUST be kept for a short time.

To handle movement, if receiving DAD NS/Gra ARP request target at the address during the period, remove the entry.

If port turns UP during the period:

?Optionally send probes to SAVI-host port for assurance;

?MUST send probes to SAVI-Poly port for assurance.

17. About Collision in Detection

The SAVI device may receive a response in detection. Some related details are specified here.

17.1. Whether to notify the DHCP server

It is unnecessary for the SAVI device to notify the DHCP server, because the host will send a DECLINE message to it once it finds the advertised address is conflict.

17.2. The result of detection without host aware

In case the SAVI device send detection packet instead of the host, the host will not be aware of the detection result. If the detection succeeds, there is no problem. However, if the detection fails, the packets from the host with the assigned address will be filtered out. This result can be regarded as a reasonable punishment for not performing duplicate detection and using a collision address.

18. Filtering during detection

In this mechanism, whenever the DHCP server replies an address, this address will be allowed immediately even before duplicate detection is completed. This design is in consideration of a host may start to send packets straightway without detection. Also this design is to be compatible with optimistic DAD [RFC4429].

However, this feature may allow an attacker to send quantities of packets with source addresses already assigned to other nodes. A practical solution for this vulnerability is configuring the address pool and allocation algorithm of the DHCP server carefully.

19. Binding Number Limitation

It is suggested to configure some mechanism in order to prevent a single node from exhausting the binding table entries on the SAVI device. Either of the following mechanism is sufficient to prevent such attack.

1. Set the upper bound of binding number for each anchor with SAVI-Host or SAVI-Poly attribute.

2. Reserve a number of binding entries for each anchor with SAVI-Host or SAVI-Poly attribute and all anchors share a pool of the other binding entries.

3. Limit DHCP Request rate per anchor, using the bound entry number of each anchor as reverse indicator.

20. MLD Consideration

The SAVI device MUST join the tentative address multicast group whenever perform duplicate detection on behavior of host.

21. Constants

MAX_DHCP_RESPONSE_TIME	120s
MAX_ARP_PREPARE_DELAY	Default 1s but configurable
MAX_ARP_DELAY	Default 1s but configurable
MAX_DAD_PREPARE_DELAY	Default 1s but configurable
MAX_DAD_DELAY	Default 1s but configurable

22. Summary of to-be-removed sections and open issues

To-be-removed sections:

1. Section 6: Conceptual data structures

2. Section 8: DHCP scenario

3. Part of Section 9: Anchor attributes

4. Section 10: Prefix configuration

Open issues (discussed but not finished):

1. Whether to keep state START

   Should the procedure be initialized based on client request or server response?

   From Eric Levy-Abegnoli and Christian Vogt.

2. Whether to keep state DETECTION

   Should DHCP interact with NDP to detect collision or should all the collision detection be left to NDP and the DHCP solution just snoop DHCP only?

   From Eric Levy-Abegnoli.

23. Security Considerations

For prefix level granularity filtering is the basis of host level granularity filtering, to learn and configure correct prefix is of great importance to this mechanism. Thus, it's important to keep RA and DHCP-PD secure. [draft-ietf-v6ops-ra-guard-03] describes a mechanism to improve the security of RA message.

24. IANA Considerations

There is no IANA consideration currently.

25. References

25.1. Normative References

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

[RFC4862] Thomson, S., Narten, T. and Jinmei, T., "IPv6 Stateless Autoconfiguration", RFC4862, September, 2007.

[RFC5227] S. Cheshire, "IPv4 Address Conflict Detection", RFC5227, July 2008.

25.2. Informative References

26. Acknowledgments

Thanks to Christian Vogt, Eric Levy-Abegnoli, Mark Williams, Erik Nordmark, Marcelo Bagnulo Braun, Alberto Garcia, Jari Arkko, David Harrington, Pekka Savola, Xing Li, Lixia Zhang, Robert Raszuk, Greg Daley, Joel M. Halpern and Tao Lin for their valuable contributions.

Authors' Addresses

Jun Bi
CERNET
Network Research Center, Tsinghua University
Beijing 100084
China
Email: junbi@cernet.edu.cn

Jianping Wu
CERNET
Computer Science, Tsinghua University
Beijing 100084
China
Email: jianping@cernet.edu.cn

Guang Yao
CERNET
Network Research Center, Tsinghua University
Beijing 100084
China
Email: yaog@netarchlab.tsinghua.edu.cn

Fred Baker
Cisco Systems
Email: fred@cisco.com