Network Working Group
Internet Draft
Intended Category: Informational
Expires: Dec. 2013

M. Sridharan
Y. Wang
P. Garg
P. Balasubramanian

June 2013

NVGRE-EXT: Network Virtualization using Generic Routing Encapsulation Extensions
draft-sridharan-virtualization-nvgre-ext-00.txt

Status of this Memo

This memo provides information for the Internet Community. It does not specify an Internet standard of any kind; instead it relies on a proposed standard. Distribution of this memo is unlimited.

Copyright Notice

This Internet-Draft will expire on March 13,6/13/201313, 2012.

Abstract

This document describes the usage of "Network Virtualization using Generic Routing Encapsulation (NVGRE)" Extensions (NVGRE-EXT). The focus of this document is on specifying the control plane operations done using NVGRE packets.

Table of Contents

1. Introduction

The NVGRE specification defines the data channel to provide network virtualization using general routing encapsulation. This document defines the control messages that are used between two network virtualization edges (NVEs) to accomplish tasks such as redirecting traffic to a new NVE when a virtual machine (VM) moves to that NVE and handling of missing policy at the NVEs.

1.1. Terminology

Please refer to [3][4][5] for more formal definition of terminology.

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 RFC-2119 [RFC2119].

In this document, these words will appear with that interpretation only when in ALL CAPS. Lower case uses of these words are not to be interpreted as carrying RFC-2119 significance.

3. NVGRE Extensions

This section describes NVGRE Extensions. NVGRE extensions are used for control messages between two NVEs to accomplish tasks such as redirection of traffic to a new NVE when a VM moves to that NVE.

The reserved VSID 0xFFFFFF is used to exchange control messages between two NVEs using the NVGRE packet format.

The following sections detail the packet format for NVGRE-EXT messages and describe the processing done by an NVE on reception of these messages.

3.1. NVGRE-EXT frame format

NVGRE header format as specified in [3] is used for communication between NVEs. NVGRE-EXT uses the reserved VSID 0xFFFFFF to send control messages.

3.2. REDIRECT message format

Let us consider a setup with 3 NVEs: sender NVE, receiver NVE and target NVE. The sender NVE is running a VM, which we will call the sender VM. The receiver NVE is running a VM, which we will call the receiver VM. The target NVE (or new NVE) is the NVE to which the receiver VM would move as a result of a VM move operation such as live migration.

Initially, the sender VM is sending packets to the receiver VM. The sender NVE encapsulates these packets in NVGRE and sends them to the receiver NVE. The receiver NVE decapsulates the packet and sends these to the receiver VM.

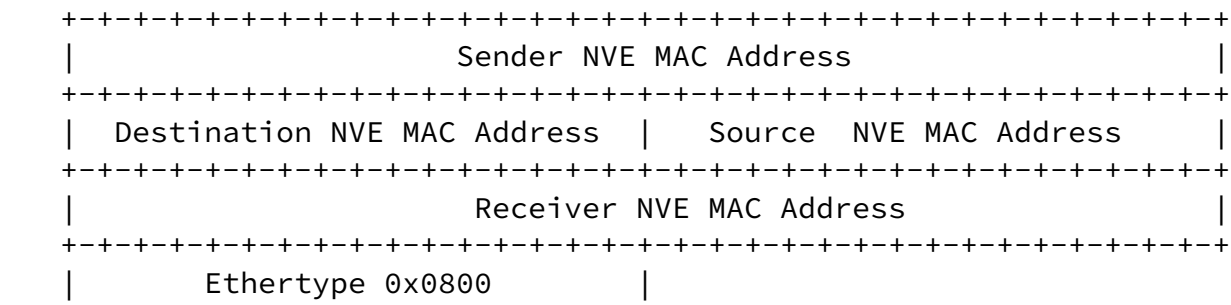
At some point, the receiver VM is moved from the receiver NVE to the target NVE. However, as the policy update may take time to reflect this movement to the other NVE, the sender NVE may continue to send the packets, from the sender VM to the receiver VM, to the receiver NVE. This can cause a connectivity interruption between the sender and receiver VMs until the policy is refreshed at the sender NVE.

The REDIRECT message is used by the receiver NVE to redirect traffic for the receiver VM to the target NVE.

This message is sent by the receiver NVE when it receives an NVGRE frame for a VM that is moved to a new NVE.

The inner Ethernet and IP frame for this message is shown in Figure 1.

Inner Ethernet Header:



Inner IP Header:

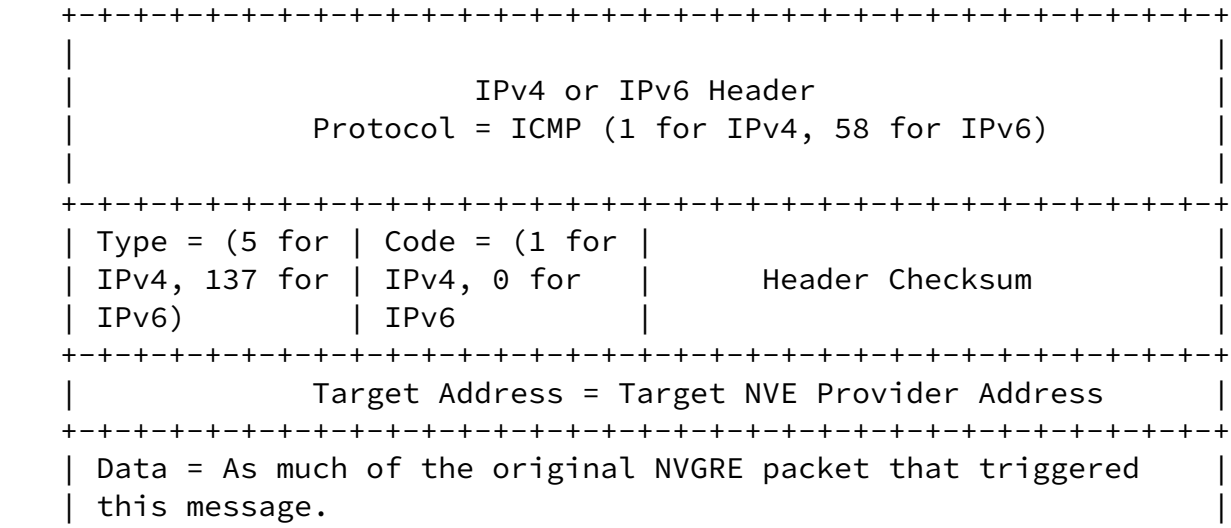


Figure 1 REDIRECT Message Format

The outer/delivery headers include the outer Ethernet header and the outer IP header:

o Ethernet header: The Ethernet header is populated with the source and destination NVE MAC addresses.

o IP header: The IP header is filled with source and destination NVE provider address respectively. The protocol field is set to ICMPv4 or ICMPv6 depending upon the IP protocol in use at NVE.

o ICMP header: The type field in the ICMP header MUST be set to 5 for IPv4 and 137 for IPv6. The code field in ICMP header MUST be set to 1 for IPv4 and 0 for IPv6.

o Target Address: The target address is set to provider address of the target NVE where the VM has moved.

o Data: The data is filled with as much of the original NVGRE frame as possible that caused this redirect message. This data MUST include the outer NVGRE frame, the inner Ethernet and IP header. This data is used by the sender NVE to identify the receiver VM that has moved to the target NVE and send further messages for that receiver VM to the target NVE.

3.3. UNREACHABLE message format

Let us consider a setup with 2 NVEs: sender NVE and receiver NVE. The sender NVE is running a VM, which we will call sender VM. The receiver NVE is running a VM, which we will call receiver VM.

Initially, the sender VM is sending packets to the receiver VM. The sender NVE encapsulates these packets in NVGRE and sends them to the receiver NVE. The receiver NVE decapsulates the packet and sends these to the receiver VM.

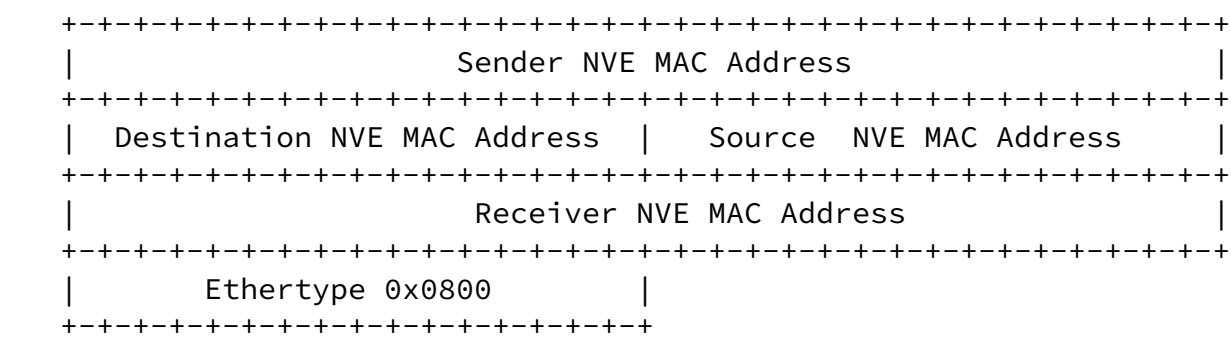
At some point, a packet is sent by the sender NVE that does not match NVGRE policy on the receiver NVE. This can cause a connectivity interruption between the sender and receiver VMs until

the NVGRE policy is refreshed at the sender NVE.

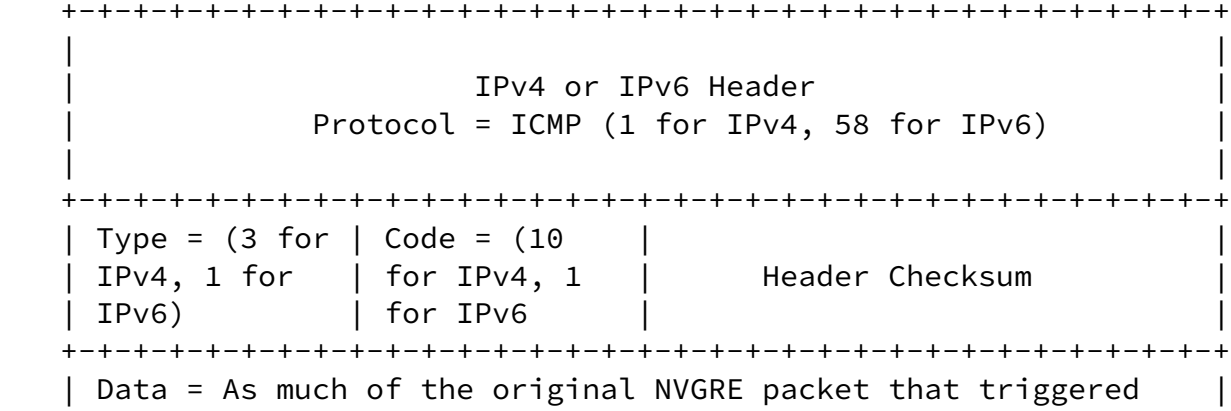
The UNREACHABLE message is used by the receiver NVE to inform the sender NVE that it needs a policy refresh. This message is sent by the receiver NVE when it receives an NVGRE frame for which it has no isolation policy.

The inner Ethernet and IP frame for this message is shown in Figure 2.

Inner Ethernet Header:



Inner IP Header:



| this message. |

Figure 2 UNREACHABLE Message Format

The outer/delivery headers include the outer Ethernet header and the outer IP header:

o Ethernet header: The Ethernet header is populated with the source and destination NVE MAC addresses.

o IP header: The IP header is filled with source and destination NVE provider address respectively. The protocol field is set to ICMPv4 or ICMPv6 depending upon the IP protocol in use at the NVE.

o ICMP header: The type field in the ICMP header MUST be set to 3 for IPv4 and 1 for IPv6. The code field in ICMP header MUST be set to 10 for IPv4 and 1 for IPv6.

o Data: The data is filled with as much of the original NVGRE frame as possible that caused this unreachable message. This data MUST include the outer NVGRE frame, the inner Ethernet and IP header. This data is used by the sender NVE to identify the receiver VM that

resulted in missing policy and refresh its NVGRE policy for the receiver VM.

4. Security Considerations

This proposal allows a faster NVGRE policy update using a REDIRECT message or an UNREACHABLE message. It can allow a compromised NVE to redirect traffic for one or more VMs. Mitigations of such attacks are possible with authentication/encryption using IPsec or any other IP based mechanism or using control plane for validation of the updated information. The control plane for NVGRE policy distribution is expected to be secured by using any of the existing security protocols and can be used to disable or override such traffic redirection decisions.

5. IANA Considerations

This document has no actions for IANA.

6. References

6.1. Normative References

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

[2] Ethertypes, ftp://ftp.isi.edu/in-notes/iana/assignments/ethernet-numbers

[3] [NVGRE] Sridharan, M., "NVGRE: Network Virtualization using Generic Routing Encapsulation", draft-sridharan-virtualization-nvgre-02, Feb 2013

6.2. Informative References

[4] M. Lasserre et al, "Framework for DC Network Virtualization", draft-ietf-nov3-framework (work in progress), February 2013.

[5] T. Narten et al, "Problem Statement: Overlays for Network Virtualization", draft-narten-nov3-overlay-problem-statement (work in progress), February 2013.

Authors' Addresses

Murari Sridharan
Microsoft Corporation
1 Microsoft Way
Redmond, WA 98052
Email: muraris@microsoft.com

Yu-Shun Wang
Microsoft Corporation
1 Microsoft Way
Redmond, WA 98052
Email: yushwang@microsoft.com

Pankaj Garg
Microsoft Corporation
1 Microsoft Way
Redmond, WA 98052
Email: pankajg@microsoft.com

Praveen Balasubramanian
Microsoft Corporation
1 Microsoft Way
Redmond, WA 98052
Email: pravb@microsoft.com