Network Working Group
Request for Comments: 4390
Category: Standards Track

Vivek Kashyap
IBM

April 2006

Dynamic Host Configuration Protocol (DHCP) over InfiniBand

Status of This Memo

Copyright Notice

Abstract

IP over Infiniband (IPoIB) link-layer address is 20 octets long. This is larger than the 16 octets reserved for the hardware address in a Dynamic Host Configuration Protocol/Bootstrap Protocol (DHCP/BOOTP) message. The above inequality imposes restrictions on the use of the DHCP message fields when used over an IPoIB network. This document describes the use of DHCP message fields when implementing DHCP over IPoIB.

Table of Contents

1. Introduction

The Dynamic Host Configuration Protocol (DHCP) provides a framework for passing configuration information to hosts on an IP network [RFC2131]. DHCP is based on the Bootstrap Protocol (BOOTP) [RFC951] adding the capability of automatic allocation of reusable network addresses and additional configuration options [RFC2131,RFC2132].

The DHCP server receives a broadcast request from a client. The DHCP server uses the client interface's hardware address to unicast a reply when the client does not yet have an IP address assigned to it. The "chaddr" field in the DHCP message carries the client's hardware address.

The "chaddr" field is 16 octets in length. The IPoIB link-layer address is 20 octets in length [RFC4391]. Therefore, the IPoIB link-layer address will not fit in the "chaddr" field making it impossible for the DHCP server to unicast a reply to the client.

To ensure interoperability, the usage of the fields and the method for DHCP interaction must be clarified. This document describes the IPoIB-specific usage of some fields of DHCP. See [RFC2131] for the mechanism of DHCP and the explanations of each field.

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 DHCP over IPoIB Mechanism

As described above, the link-layer address is unavailable to the DHCP server because the link-layer address is larger than the "chaddr" field length. As a result, the server cannot unicast its reply to the client. Therefore, a DHCP client MUST request that the server send a broadcast reply by setting the BROADCAST flag when IPoIB Address Resolution Protocol (ARP) is not possible, i.e., in situations where the client does not know its IP address.

[RFC1542] discourages the use of a broadcast reply. But in the case of IPoIB, this is a necessity because the server does not receive the link-layer address. To desynchronise broadcasts at subnet startup, [RFC2131] suggests that a client wait a random time (1 to 10 seconds) before initiating server discovery. The same timeout will spread out the DHCP server broadcast responses generated due to the use of the BROADCAST bit.

The client hardware address, "chaddr", is unique in the subnet and hence can be used to identify a client interface. But in the absence of a unique "chaddr", another unique client identifier must be used.

The DHCP protocol states that the "client identifier" option may be used as the unique identifying value for the client [RFC2132]. This value must be unique within the client's subnet.

The "client identifier" option includes a type and identifier pair. The identifier included in the "client identifier" option may consist of a hardware address or any other unique value such as the DNS name of the client. When a hardware address is used, the type field should be one of the ARP hardware types listed in [ARPPARAM].

2.1. IPoIB-specific Usage of DHCP Message Fields

A DHCP client, when working over an IPoIB interface, MUST follow the following rules:

"htype" (hardware address type) MUST be 32 [ARPPARAM].

"hlen" (hardware address length) MUST be 0.

"chaddr" (client hardware address) field MUST be zeroed.

"client-identifier" option MUST be used in DHCP messages.

The "client identifier" used in DHCP messages MUST conform to [RFC4361].

2.2. Use of the BROADCAST flag

A DHCP client on IPoIB MUST set the BROADCAST flag in DHCPDISCOVER and DHCPREQUEST messages (and set "ciaddr" to zero) to ensure that the server (or the relay agent) broadcasts its reply to the client.

Note: As described in [RFC2131], "ciaddr" MUST be filled in with the client's IP address during BOUND, RENEWING or REBINDING states; therefore, the BROADCAST flag MUST NOT be set. In these cases, the DHCP server unicasts DHCPACK message to the address in "ciaddr". The link address will be resolved by ARP.

3. Security Considerations

[RFC2131] describes the security considerations relevant to DHCP. This document does not introduce any new issues.

4. Acknowledgement

This document borrows extensively from [RFC2855]. Roy Larsen pointed out the length discrepancy between the IPoIB link address and DHCP's "chaddr" field.

5. References

5.1. Normative References

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

[RFC2131] Droms, R., "Dynamic Host Configuration Protocol", RFC 2131, March 1997.

[RFC2132] Alexander, S. and R. Droms, "DHCP Options and BOOTP Vendor Extensions", RFC 2132, March 1997.

[RFC951] Housley, R., Horting, T., and P. Yee, "TELNET Authentication Using KEA and SKIPJACK", RFC 2951, September 2000.

[RFC4391] Chu, J. and V. Kashyap "Transmission of IP over InfiniBand (IPoIB)", RFC 4391, April 2006.

[ARPPARAM] http://www.iana.org/numbers.html

[RFC4361] Lemon, T. and B. Sommerfeld, "Node-specific Client Identifiers for Dynamic Host Configuration Protocol Version Four (DHCPv4)", RFC 4361, February 2006.

5.2. Informative References

[RFC2855] Fujisawa, K., "DHCP for IEEE 1394", RFC 2855, June 2000.

[RFC1542] Wimer, W., "Clarifications and Extensions for the Bootstrap Protocol", RFC 1542, October 1993.

Author's Address

Vivek Kashyap
15350, SW Koll Parkway
Beaverton, OR 97006
USA

Phone: +1 503 578 3422
EMail: vivk@us.ibm.com

Full Copyright Statement

Intellectual Property

Acknowledgement

Funding for the RFC Editor function is provided by the IETF Administrative Support Activity (IASA).