INTERNET DRAFT
<draft-ietf-ipoib-dhcp-over-infiniband-10.txt>
Expiration Date: September 2005

Vivek Kashyap
IBM

March 2005

DHCP over InfiniBand

Status of this memo

By submitting this Internet-Draft, I certify that any applicable patent or other IPR claims of which I am aware have been disclosed, or will be disclosed, and any of which I become aware will be disclosed, in accordance with RFC 3668.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

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."

The list of current Internet-Drafts can be accessed at http://www.ietf.org/ietf/1id-abstracts.txt.

The list of Internet-Draft Shadow Directories can be accessed at http://www.ietf.org/shadow.html.

Copyright (C) The Internet Society (2001). All Rights Reserved.

Abstract

An InfiniBand network uses a link-layer addressing scheme that is 20-octets long. This is larger than the 16-octets reserved for the hardware address in DHCP/BOOTP message. The above inequality imposes restrictions on the use of the DHCP message fields when used over an IP over InfiniBand (IPoIB) network. This document describes the use of DHCP message fields when implementing DHCP over IPoIB.

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 the DHCP client. The DHCP server uses the client interface's hardware-address to unicast a reply back when the client doesn't 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. Therefore the IPoIB link-layer address will not fit in the "chaddr" field making it impossible for the DHCP server to unicast a reply back 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 back to the client. Therefore, a DHCP client MUST request that the server sends a broadcast reply by setting the BROADCAST flag when IPoIB ARP is not possible, i.e. in situations where the client does not know its IP address.

[RFC1542] notes that the use of a broadcast reply is discouraged. 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 equally spread out the DHCP server broadcast responses generated due to the use of the use of the BROADCAST bit.

The client hardware address, "chaddr", is unique in the subnet and hence can be used to identify the 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 subnet the client is a member of.

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 [DHC_3315id].

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 client's IP address during BOUND, RENEWING or REBINDING state, 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

[RFC2119] Key words for use in RFCs to Indicate Requirement Levels, S. Bradner

[RFC2131] Dynamic Host Configuration Protocol, R. Droms

[RFC2132] DHCP Options and BOOTP Vendor Extensions, S. Alexander, R. Droms

[RFC951] Bootstrap Protocol, B. Croft, J. Gilmore

[IPoIB_ENCAP] draft-ietf-ipoib-ip-over-infiniband-09.txt, H.K. Jerry Chu, V. Kashyap

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

[IBARCH] InfiniBand Architecture Specification, www.infinibandta.org/specs

[IPoIB_ARCH] draft-ietf-ipoib-architecture-04.txt, V. Kashyap

[DHC_3315id] draft-ietf-dhc-3315id-for-v4-04.txt, T. Lemon, B. Sommerfeld

5.2 Informative

[RFC2855] DHCP for IEEE 1394, K. Fujisawa

[RFC1542] Clarifications and Extensions for the Bootstrap Protocol, W. Wimer

6. 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

Copyright (C) The Internet Society (2004). This document is subject to the rights, licenses and restrictions contained in BCP 78, and except as set forth therein, the authors retain all their rights.

This document and the information contained herein are provided on an "AS IS" basis and THE CONTRIBUTOR, THE ORGANIZATION HE/SHE REPRESENTS OR IS SPONSORED BY (IF ANY), THE INTERNET SOCIETY AND THE INTERNET ENGINEERING TASK FORCE DISCLAIM ALL WARRANTIES, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO ANY WARRANTY THAT THE USE OF THE INFORMATION HEREIN WILL NOT INFRINGE ANY RIGHTS OR ANY IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

Intellectual Property Statement

The IETF takes no position regarding the validity or scope of any Intellectual Property Rights or other rights that might be claimed to pertain to the implementation or use of the technology described in this document or the extent to which any license under such rights might or might not be available; nor does it represent that it has made any independent effort to identify any such rights. Information on the procedures with respect to rights in RFC documents can be found in BCP 78 and BCP 79.

Copies of IPR disclosures made to the IETF Secretariat and any assurances of licenses to be made available, or the result of an attempt made to obtain a general license or permission for the use of such proprietary rights by implementers or users of this specification can be obtained from the IETF on-line IPR repository at http://www.ietf.org/ipr.

The IETF invites any interested party to bring to its attention any copyrights, patents or patent applications, or other proprietary rights that may cover technology that may be required to implement this standard. Please address the information to the IETF at ietf-ipr@ietf.org.

Acknowledgment

Funding for the RFC Editor function is currently provided by the Internet Society.